AD-A181 697

AAMRL-TR-87-007
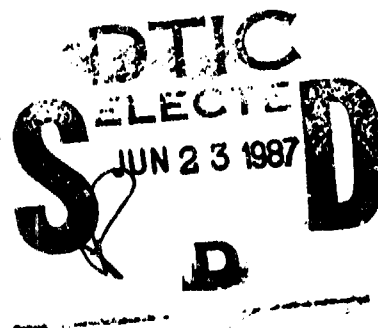
UNIFIED TRI-SERVICES COGNITIVE PERFORMANCE
ASSESSMENT BATTERY: REVIEW AND METHODOLOGY (U)

WILLAM A. PEREZ
PHILIP J. MASLINE
ERIC G. RAMSEY
KENNETH E. URBAN

SYSTEMS RESEARCH LABORATORIES, INC.

DTIC
SELECTED
JUN 23 1987
S D
D

MARCH 1987

Approved for public release; distribution is unlimited.

87   6 22   079

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
UNCLASSIFIED	

2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
	Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
	AAMRL-TR-87-007

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
Systems Research Laboratories, Inc.		AAMRL/HEG

6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
2800 Indian Ripple Road Dayton, Ohio 45440-3696	Wright-Patterson AFB OH 45433-6573

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
U.S. Army Medical Research & Development Command		F33615-85-C-0541

8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO
Fort Detrick, Maryland 21701-5012	63764A	2729	10	01

11. TITLE (Include Security Classification)

Unified Tri-Services Cognitive Performance Assessment Battery: Review and Methodology (U)

12. PERSONAL AUTHOR(S)
Perez, William A., Masline, Philip J., Ramsey, Eric G. and Urban, Kenneth E.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM Apr 84 TO Feb 87	1987 March	342

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Human Performance, Task Battery
			Human Performance Tests

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The Unified Tri-Services Cognitive Performance Assessment Battery (UTC-PAB) represents the primary metric for a Level 2 evaluation of cognitive performance in the JWGD3 MILPERF chemical defense biomedical drug screening program. The UTC-PAB contains a menu of 25 tests that were selected from test batteries currently in existence throughout DoD research laboratories. Test selection was based upon established test validity and relevance to military performance. Sensitivity to effects of hostile environments and sustained operations were also considerations involved in test selection.

This report presents a scheme for organizing the tests in the UTC-PAB. Also, extensive documentation for each test is presented in the following areas: background literature review focusing on the theoretical basis of the test; information regarding the reliability, validity, and sensitivity of the test; data specifications; and instructions to subjects. This information is presented to guide researchers in the selection and interpretation of tests in the UTC-PAB. (Keywords:

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Ronald E. Yates	(513) 255-7587	AAMRL/HEG

DD FORM 1473, 84 MAR
83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

SUMMARY

The Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB) is the primary instrument for the assessment of cognitive performance in a multiple level drug evaluation program (The Military Performance Working Group, 1983). The UTC-PAB consists of a computerized test system (see Hegge et al., 1985 for a description) and supporting documentation. The present report provides literature reviews and sections on methodology for each of the 25 tests that were selected by the Tri-Service Joint Working Group on Drug Dependent Degradation of Military Performance (JWGD3 MILPERF). The report by Englund et al. (1985) presents the historical overview of UTC-PAB construction, the rationale, and criteria for test selection and a framework by which to organize the 25 tests.

This report presents the organizational scheme that was proposed by Englund et al. (1985). In addition, the following sections are provided for each test: (a) Purpose, (b) Description, (c) Background, (d) Reliability, (e) Validity, (f) Sensitivity, (g) Technical Description, (h) Trial Specification, (i) Data Specification, (j) Training Requirements, and (k) Instructions to Subjects. The organization scheme and the detailed information on each test can be used to select tests that meet specific research requirements.

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

PREFACE

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

BLANK PAGE

Section 1
INTRODUCTION

The Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB) is the primary instrument for the assessment of cognitive performance in a multiple level drug evaluation program (The Military Performance Working Group, 1983). Figure 1 illustrates the relationship between the UTC-PAB and the entire drug testing program. The UTC-PAB is one of the test instruments that will be used during the level 2 testing phase. Figure 2 shows the relationship between the UTC-PAB and other test instruments to be used during level 2 drug testing (Perez, 1985). In addition, this figure illustrates the fact that the UTC-PAB will consist of a computerized test system and supporting documentation (Hegge et al., 1985). The present document presents the 25 tests that were selected by the Tri-Service Joint Working Group on Drug Dependent Degradation of Military Performance (JWGD3 MILPERF). The report by Englund et al. (1985) presents the historical overview of UTC-PAB construction, the rationale and criteria for test selection and a framework by which to organize the 25 tests. The framework proposed in the above report will be presented in this document; however, the reader is advised to read Englund et al. (1985) for information regarding the formulation of the UTC-PAB.

The framework that was selected is based on two dimensions that are particularly critical to the assessment of drug effects on cognitive performance: (a) the stage of information processing which is most markedly affected by the demands of the task, and (b) the requirement to divide or selectively employ attentional capacity between sources of information. Several major functions can be distinguished within the stages of processing dimensions. These include perceptual input functions, such as information detection and identification; central processing functions, including a variety of memory and information integration/manipulation functions; and, motor output or response execution functions (Shingledecker, 1984). Integration and manipulation functions within central processing can be further subdivided into those based on symbolic/linguistic forms of information versus those involving spatial information.

7

CIVILIAN TECH BASE
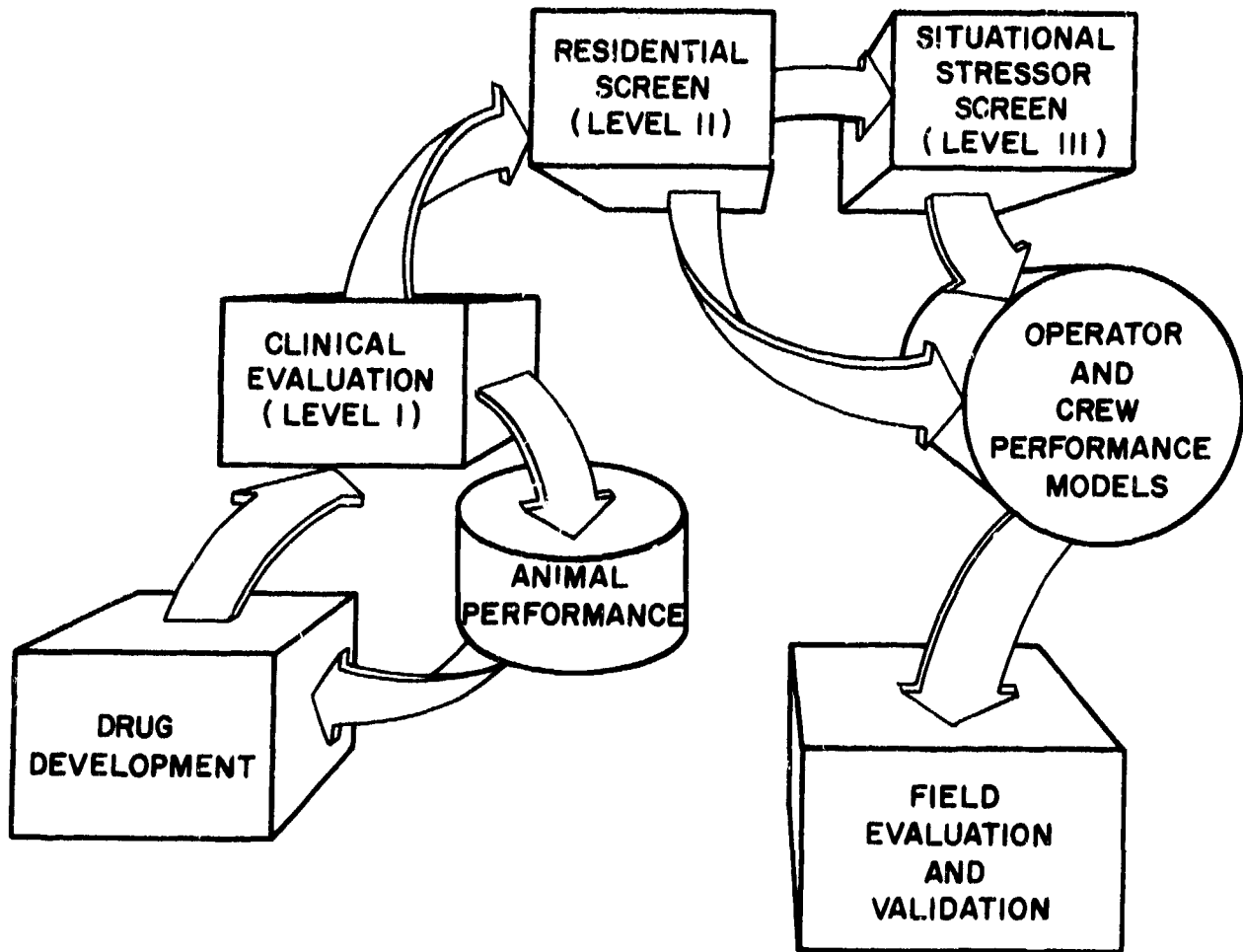
MILITARY REQUIREMENTS

Figure 1. The Tri-Service Drug Screening Program
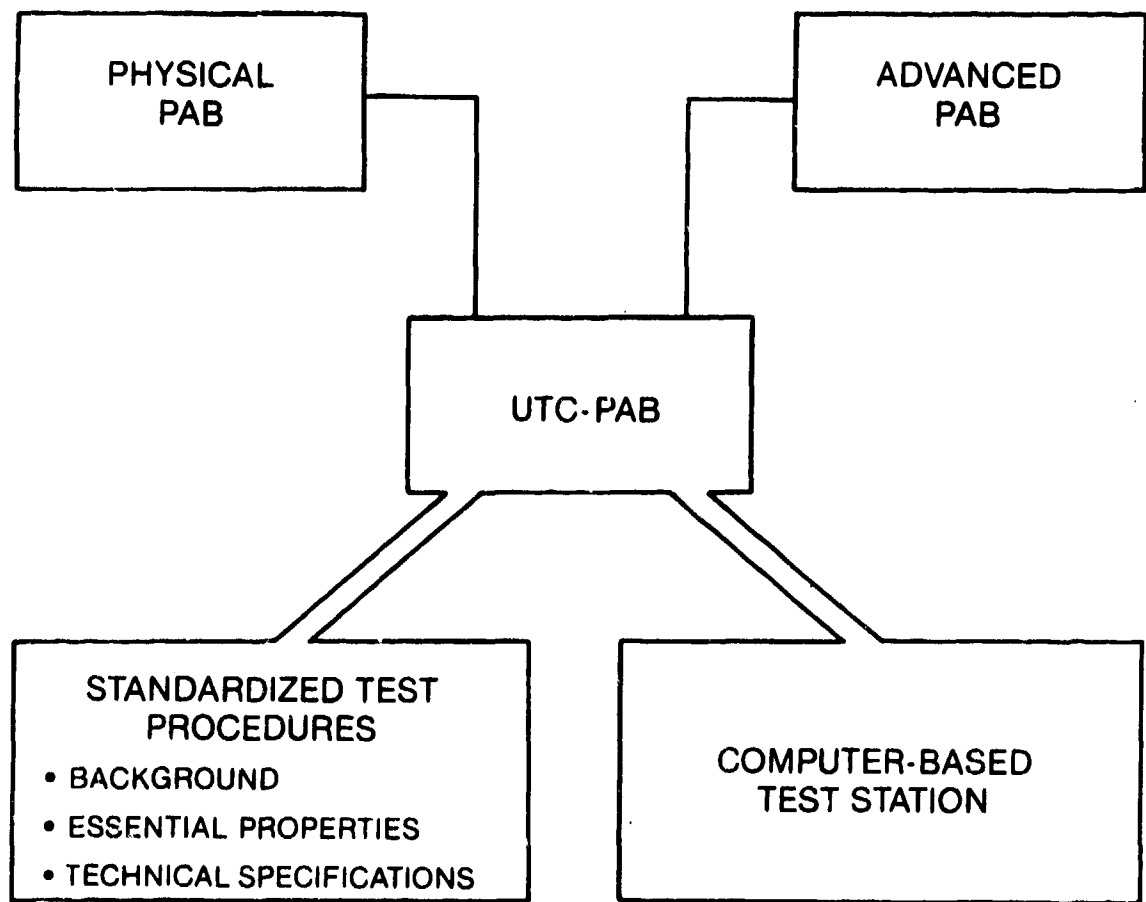
Figure 2.   The UTC-PAB Computer Based Test Station and Standardized
Test Procedures in Relation to Other Level 2 Test Systems

9

Table 1 presents the framework presented by Englund et al. (1985) for organizing the tests within the UTC-PAB. This framework was presented as a guideline for selecting subsets of tests from the UTC-PAB for particular applications. For example, one typical use of the battery would consist of an initial overall screening of the effects of a drug on major information processing functions, followed by a more extensive and diagnostic evaluation of those functions which proved to be degraded during the initial screening. In most applications, it is desirable that an overall or global screening be conducted with a subset of tasks that are representative of the major processing functions listed under Table 1. The following is one example of a subset of tests that could be used in an initial screen:

EXAMPLE OF AN INITIAL SCREEN

- Memory Search
- Mathematical Processing
- Successive Pattern Comparison
- Unstable Tracking
- Memory Search/Unstable Tracking Combination

The above subset is one of several options that would represent the various stages of processing functions included in the framework. Future research with the UTC-PAB may result in the formulation of a core subset of tests to be used for the evaluation of drug effects on cognitive performance; however, such a core set of tests cannot be recommended at this time due to the lack of empirical data.

Depending upon the pattern of results from the initial global screening, particular functions could be selected for further investigation. For example, if the global evaluation outlined above indicated that the Memory Search and Mathematical Processing tests were principally affected by a particular drug, the memory and symbolic information manipulation/integration functions would represent inportant candidates for more extensive and diagnostic investigation. This investigation would be accomplished through the choice of additional subsets of tests from the memory and symbolic information manipulation components of the UTC-PAB.

10

TABLE 1. UTC-PAB ORGANIZATION SCHEME

I. PERCEPTUAL INPUT, DETECTION, AND IDENTIFICATION

- Visual Scanning Task (16)
- Visual Probability Monitoring Task (18)
- Pattern Comparison (Simultaneous) (14)
- Four-Choice Serial Reaction Time (8)

II. CENTRAL PROCESSING

- Auditory Memory Search (Memory Search Tasks) (10)
- Continuous Recognition Task (7)
- Code Substitution Task (17)
- Visual Memory Search (Memory Search Tasks) (10)
- Item Order Test (26)

III. INFORMATION INTEGRATION/MANIPULATION--LINGUISTIC/SYMBOLIC

- Linguistic Processing Task (2)
- Two-Column Addition (5)
- Grammatical Reasoning (Symbolic) (4)
- Mathematical Processing Task (6)
- Grammatical Reasoning (Traditional) (3)

IV. INFORMATION INTEGRATION/MANIPULATION--SPATIAL MODE

- Spatial Processing Task (11)
- Matching To Sample (25)
- Time Wall (19)
- Matrix Rotation Task (Spatial Processing Task) (11)
- Manikin Test (13)
- Pattern Comparison (Successive) (15)

V. OUTPUT/RESPONSE EXECUTION

- Interval Production Task (20)
- Unstable Tracking Task (23)

VI. SELECTIVE/DIVIDED ATTENTION

- Dichotic Listening Task (22)
- Memory Search/Unstable Tracking Combination (24) (Sternberg-Tracking Combination)
- Stroop Test (21)

NOTE: The number following the test name corresponds to the sections in this report.

The present report provides extensive documentation for each test in the UTC-PAB to aid in the selection and interpretation of test results. The following sections are included for each test:

- Purpose — A brief statement indicating the cognitive function which the test evaluates (e.g., working memory, motor response processing, etc.).

- Description — A nontechnical description of the test which outlines the subjects' task.

- Background — A thorough literature review of the test.

- Reliability — Information pertaining to test-retest reliability.

- Validity — This section focuses on a test's construct validity.

- Sensitivity — Information regarding the uses of UTC-PAB tests (or equivalent versions) in the areas of behavioral toxicology, behavioral drug testing, environmental stress research.

- Technical Description — A description of the test with sufficient details for the development of computer programs.

- Trial Specifications — A step by step description of each trial in a test.

- Data Specification — The nature of the data generated by a test. In addition, cautionary statements with respect to parametric properties or violations are provided when needed.

- Training Requirements — If possible, information indicating the number of trials required to reach stable levels of performance are presented. However, this type of information is not available for many of the tests in the UTC-PAB. In addition, recommended procedures for familiarizing subjects with the tests are presented.

- Instructions To Subjects — Detailed instructions to subjects are provided. It is important to standardize the instructions to subjects since significant variations in responses can be obtained by varying instructions (e.g., vary speed accuracy requirements).

It should be noted that the tests in the UTC-PAB were selected from test batteries that had been in existence within DoD for some time. These original test batteries are still in use within the DoD research community and are undergoing revisions. For example, the Unstable Tracking,

12

Continuous Recognition, and Probability Monitoring tests have undergone significant revisions after the specifications for these tests were submitted to the JWGD3 for inclusion in the UTC-PAB. The above modified tests represent significant improvements relative to the versions that were originally included in the UTC-PAB. However, these modified test versions were just recently validated and we were unable to include them in our present documentation of the UTC-PAB. Information regarding these modified tests is presented in Appendix A to this report.

This document represents an initial effort to integrate and standardize cognitive performance assessment for the screening of chemical defense treatment and pretreatment drugs. The UTC-PAB represents a "menu" of tests from which to select those tests that meet specific research requirements. The organization scheme that was presented earlier can be used as a guideline for selecting tests; however, this is just one of many different organizational schemes that could be proposed and should not be interpreted as the "model" for the UTC-PAB. Documentation for the UTC-PAB should be an ongoing effort that incorporates the results of the JWGD3 drug evaluation program. Tests that are currently in the battery may be modified or deleted and new tests may be introduced to meet the demands of the drug testing program (e.g., additional tests that address selective/divided attention).

LINGUISTIC PROCESSING TASK (UTC-PAB TEST NO. 1)
(VISUAL AND SEMANTIC CODING)

PURPOSE

The purpose of the Linguistic Processing Task is to test a subject's ability to code linguistic information at different depths of processing. The task places variable demands upon the resources associated with the processing and transformation of linguistic information.

DESCRIPTION

This task is a synthesis of Posner and Mitchell's (1967) letter match task and generic depth of processing tasks (e.g., Craik and Tulving, 1975). It is a standardized loading task which requires classification of letter or word pairs. Letter or word pairs are presented on a CRT, and subjects are instructed to respond "same" if the items match on the dimension in question or "different" if otherwise. Three levels of task demand are imposed by the following classification rules: Physical letter match, in which letter pairs must be physically identical to match (low demand); category match, requiring that both letters are either consonants or vowels (moderate demand); and antonym match, in which only words opposite in meaning constitute a match (high demand). Each set of trials lasts three minutes.

BACKGROUND

Posner and Mitchell (1967) designed an experiment that provided an opportunity to observe processing at different levels within the same paradigm. The goal of the study was to find levels of processing that depend primarily upon the physical attributes of the stimulus and levels which depend upon a more detailed analyses such as naming or relating to a subordinate. In the experiment, the stimuli were pairs of letters, digits, or forms and the subject was always pressing one of two keys ("same" or "different"). The subjects were instructed to classify the stimulus pair based upon some predetermined rule. There were three different levels of

classification rules. The instructions used to define "same" were physical identity (e.g., AA), name identity (e.g., Aa), or rule identity (e.g., both vowels). The experiment was designed to determine if the different levels of instruction produced orderly differences in the rate at which subjects made the classification.

Pairs of capital and small case letters were visually presented simultaneously to the subjects. The subject then classified the letter pair based upon one of the three rules. The letters remained present until the subject responded by pressing a switch. Level 1 instructions were to classify each pair of stimuli "same" if they were physically identical and "different" if they were not. Level 2 instructions were to classify letters "same" if they had the same name and "different" if they did not. Level 3 instructions were to classify letters "same" if they were both vowels or both consonants and "different" if they were mixed. The subjects were instructed to classify each pair as rapidly as possible, trying to keep errors to a minimum. Reaction times from stimulus onset until response were recorded.

The results showed a significant effect of classification rule. Different instructions led to significant differences in mean RT. A second experiment (directly comparing levels 1 and 2) demonstrated a significant difference in mean RT, with level 1 RTs shorter. Based on the obtained RTs, the authors infer three different processing nodes. The first is based on physical identity and includes letter pairs that are identical in form. This type of match is believed to be free of prior learning effects. The second node is based on name identity. This involves matching letters which have no obvious physical similarity so that the subject must derive something like the name of the letter in order to make the match. Since matches based on a common name were found to be reliably faster than those based on a common rule (vowel-vowel or consonant-consonant), rule identity was considered as a third node or level of processing.

The depth of processing framework for human memory research was expanded on in a series of experiments by Craik and Tulving (1975). Depth of processing here refers to greater degrees of semantic involvement. Subjects were

15

induced to process words to different depths by answering various questions about the words. For example, shallow encodings were achieved by asking questions about typescript; intermediate levels of encoding were accomplished by asking questions about rhymes; deep levels were induced by asking whether the word would fit into a given category or sentence frame. After the encoding phase was completed, subjects were unexpectedly given a recall or recognition test for the words. In general, deeper encodings took longer to accomplish and were associated with higher levels of performance on the subsequent memory test. Also, questions leading to positive responses were associated with higher retention levels than questions leading to negative responses, at least at deeper levels of encoding.

In the experiment, a different word was exposed on every trial. Before the word was exposed, the subject was asked a question about the word. Three types of questions were asked: (1) An analysis of the physical structure of the word was affected by asking questions such as "Is the word printed in capital letters?" (2) A phonemic level of analysis was induced by asking about the words rhyming characteristics. (3) A semantic analysis was activated by asking categorical questions (e.g., Is the word an animal name?).

Results showed that response latency rose systematically as the question necessitated deeper processing. Questions about the surface form of the word were answered comparatively rapidly, while more abstract questions about the word took longer to answer. Same responses took 591, 614, and 689 milliseconds (msec) for physical, name, and category matches respectively. No significant differences between same and different responses were found. This research provided further support for the notion that memory performance depends on the depth to which the stimulus is analyzed.

Subsequent studies involving the linguistic processing task have examined the manipulations of various stimulus variables on encoding times and depths of processing. A few of these studies will now be described. An experiment conducted by Posner et al. (1969) varied the match type (physical same, name same, different), and the (ISI) interstimulus interval (0, .5, 1, or 2 seconds) in a letter match paradigm. Reaction times were recorded as the dependent measure. The results showed a significant

interaction between the same match types and ISI. The difference in reaction time between physical and name matches decreases with increases in ISI. Posner et al. (1969) concluded that matches based on visual information (physical) becomes relatively less efficient over time, possibly because: (1) the visual code loses clarity, (2) visual cues lose saliency over time, or (3) the name information becomes more efficient.

Judgements of same typically have a shorter response time than judgements of different (Krueger, 1978). Also, when subjects are required to match on the basis of name, the judgements that the target stimuli have the same name is more rapid when the stimuli are physically identical than when one of the targets is the upper--and the other is the lowercase version of the letter. This difference in latencies between same and different judgements is attributed to response competition between name codes. The response competition model of simultaneous matching tasks attributes the longer latency for different judgements to a greater degree of response competition when the stimuli to be matched are different. Response competition was found to be a significant factor in determining differences in latency for same/different responses to physical matches (Eriksen, O'Hara, and Eriksen, 1982). This was not proved, however, for name matches (Eriksen and O'Hara, 1982).

Many experiments involving the letter match task have focused on the differences in reaction time between physical and name matches. For example, Kirsner, Wells, and Sang (1982) examined the effects of different typefonts on RT in a letter match task. In the study, RT was found to decrease with increasing similarity of font. Visual as well as acoustic confusability has also been tested by Thorson, Hochhaus, and Stanners (1976). In this letter matching task, letter pairs were presented that were either visually confusable, acoustically confusable, or both. The effects on RT were examined. Results suggested that visual coding is emphasized for approximately 1 second, after which acoustic code seems to dominate.

In some versions of linguistic processing tasks, words are matched instead of letters. Marmurek (1977) investigated the differences in processing between words and letters in this type of task. In the study, subjects

17

indicated whether two letters, two words, or a letter and the first letter of a word were the same. Letter targets were matched more quickly than word targets when the stimuli were presented simultaneously. However, when a 3-second interval separates target and comparison presentation, word targets are matched more quickly than a letter and a letter in a word. These findings support a level of processing model of word processing. Identification of letters occurs at a "lower" level of processing, while an entire word can be encoded as a unit at a "higher" (more elapsed time) level of processing. Words, however, take longer to process than letters regardless of the classification rule imposed.

Both words and letters were matched in a version of the linguistic processing task developed by Shingledecker (1984). This task combines letter matching tasks (e.g., Posner and Mitchell, 1967) with depths of processing tasks (e.g., Craik and Tulving, 1975). Three significantly different demand conditions are imposed by the following classification rules: physical letter match in which letter pairs must be physically identical to match (low demand); category match requires that both letters be either consonants or vowels (moderate demand); and, antonym match in which only words opposite in meaning constitute a match (high demand). These conditions have been shown to place variable demands upon mental resources associated with the manipulation and comparison of linguistic information.

The UTC-PAB version of the linguistic processing task is identical to that of Shingledecker (1984) described above. This task utilizes the physical and category classification rules as found in Posner and Mitchell (1967) but not the name match. Although significant differences in response time have been determined between physical and name matches, experimenters do not agree that visual and phonetic coding involve independent processing and depths of processing. Category match and antonym match have never been compared in the same experiment except for the Shingledecker (1984) study. Processing of words has been shown to be a higher level than letters and is accompanied by longer response times (Marmurek, 1977). Also, determining the antonym of a word requires higher level thought (deeper processing) than determining the relationship of two letters as vowels or consonants.

18

RELIABILITY

No reliability studies (e.g. test-retest) have been performed on the
current version of the UTC-PAB linguistic processing task. However, a
reliability study involving the three levels of processing of the Posner
letter matching task has been conducted and will now be described.

Harbeson, Kennedy, Krause, and Bittner (1982) performed a repeated measures
analysis of Posner's letter matching test for its inclusion in the Perform-
ance Evaluation Tests for Environmental Research (PETER) battery. In the
experiment, 21 subjects were tested for 15 minutes per day for 15 consecu-
tive work days. Subjects were to make same or different judgements on
pairs of letters based on three criteria. Letters were classified by phy-
sical appearance (AA versus AB), name identity (Aa versus Ab), or category
(both vowels or consonants such as AE or BC versus not matched, such as
AB). There were 36 trials per day in each of the first two conditions and
32 in the third. The number of trials was sufficient to observe means at
asymptote, and provided sufficient data for tests of the stability of var-
iances and correlations. The interstimulus interval was approximately
4 seconds. Dependent measures included response times for each condition
for same judgements, response times for all different judgements, two
difference scores, percent errors, and mean error times. Means, standard
deviations, and cross session correlations were calculated for each
measure.

Response times to the task stabilized after 8, 10, and 12 days for name,
physical, and category matches respectively. Reliability coefficients were
.81, .83, and .89 for physical, name, and category matches respectively.
All three measures were also very highly correlated (physical-name, .99;
physical-category, .90; name-category, .94). The authors concluded that
since these measures appear to be redundant within tests, the Posner letter
matching task would be suitable for repeated measures (environmental)
testing.

VALIDITY

The linguistic processing task does seem to test a person's ability to encode information at different levels of processing. The finding that different levels of processing are defined by the physical, naming, and categorical classification of a stimulus has been well established in numerous letter match investigations. The level of processing framework developed by Craik and Tulving (1975) has also established itself as a valid approach to explaining memory processes. In all studies, the higher the level of encoding of the stimulus (more deeply processed) the longer the response time to a comparison of the words or letters. The physical classification of stimuli and the classification of both vowels or both consonants versus one vowel and one consonant have been validated within the same experimental paradigm. Antonym matches have not been used in this type of task to any great extent and their relation to the other two levels has not been established.

SENSITIVITY

The linguistic processing task has not been used in studies investigating the effects of environmental stressors. However, the reliability and validity of the test, as well as the levels of processing model, provide a framework for deriving predictions with respect to the effect of stressors on cognitive processing. Performance on the task should break down as a function of the level of processing. That is, under environmental stress deeper levels of processing (antonym matching) would be predicted to be interfered with first. As more stress is experienced, the performance of lower levels of processing should also deteriorate (category match and then physical match). Investigations supporting these predictions are lacking at this time.

TECHNICAL DESCRIPTION

Letter pairs to be presented for the physical identity and category match rules are selected from the population of all possible (64) combinations of both upper and lower case versions of the letters A, B, C, and E. Same and

different letter pairs are randomly generated with equal probability. Antonyms were taken from Roget's Thesaurus. Individual words composing the antonyms are paired with both matching and nonmatching words throughout testing. Letters presented are approximately .5 by .7 cm and are viewed from a distance of roughly 62 cm. Same and different responses are entered on appropriately labeled keys.

A maximum response time or "deadline" is imposed in each condition. Stimuli are displayed until the subject responds or until the deadline is reached, thus allowing subjects to pace themselves within the restrictions imposed by the deadline. During training, the deadline is set at 15 seconds for all conditions. More restrictive deadlines are used on testing trials. For the physical identity match condition, the testing deadline is 1 second; for the category match condition, 1.5 seconds; and for the antonym match conditon, 1.5 seconds. Subjects are instructed before each set of trials as to which classification rule (physical, category, or antonym) they will be using for that trial. Each set of trials lasts 3 minutes.

DATA SPECIFICATIONS

Unprocessed data are collected and stored on all trials. These data will be a record of: (1) trial start time, (2) problem onset time, (3) subject response (match or nonmatch), and (4) response latency in msec.

From these raw data measurements the following summary statistics can be computed for each trial: (1) number of problems presented, (2) number and percent correct responses, (3) total percent errors, (4) percent errors of omission (failure to respond before deadline), (5) percent errors of commission (incorrect response), (6) mean and median correct response time, and (7) standard deviation of response time.

TRAINING REQUIREMENTS

Depending upon the condition being tested, trials begin by giving subjects the appropriate rule to be used in determining whether or not the letter or

word pairs constitute a match (physical identity of the stimulus letters, both vowels or both consonants, or opposite meaning of words). Subjects are told to respond as quickly and accurately as possible. Major practice effects are attenuated with five to 10 3-minute training trials at each loading level (Shingledecker, 1984).

To summarize, the training phase for this test should consist of the following steps:

1.   Read instructions to the subjects.

2.   Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3.   Repeat the practice trials if it appears that the subjects require additional practice with the test.

4.   Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

This task requires you to classify pairs of letters or words as "same" or "different" on the basis of their shape, grammatical category, or meaning. In one level of the task, pairs of upper or lower case versions of the letters A, B, C, and E are presented one at a time on the screen, and you are to decide whether the two letters are physically identical. If the stimulus pair AA was presented, you would respond by pressing the key labeled "same," since the two letters have exactly the same shape. If you saw Aa you would respond to the "different" key. Although both letters are As, they have a different shape. This level of the task is called the "physical identity match."

Another level of the task is called the "category match." Pairs of upper and lower case versions of the letters A, B, C, and E are again shown one at a time, and you must decide whether both of the letters are vowels or both consonants ("same") or whether one letter is a vowel and the other is a consonant ("different"). As an example, EC would be "different," since E is a vowel and C is a consonant. Bc would be "same" because both B and C are consonants.

The third level of the task is known as the "antonym match." In this condition, pairs of words are presented together on the screen, and you must decide whether the words are opposite in meaning ("same") or not ("different"). For example, the words LAWFUL-CRIMINAL have the opposite meaning, and you should respond "same." ETERNAL-NONSENSE are not opposite in meaning, so a "different" response would be correct.

The task is performed in 3-minute trial periods. You start the data collection when you are ready by pressing either of the response keys. Stimuli will appear one pair at a time, and you should attempt to respond as quickly and accurately as possible. As soon as you enter a response, the next problem will appear. Respond as quickly as you can when answering each item, but if you find yourself making errors, slow down. You should try to get every item right. Three minutes after you press the response key to start the trial, the task will automatically stop and the screen will go blank.

Section 3
GRAMMATICAL REASONING (TRADITIONAL) (UTC-PAB TEST NO. 2)
(LOGICAL REASONING)

PURPOSE

The purpose of the grammatical reasoning test is to measure the subject's general reasoning ability. This test is a type of sentence verification task that taps the processing capacity of working memory. Furthermore, it is known to be sensitive to environmental stress, pollutants, and the effects of sleep loss.

DESCRIPTION

During this test, pairs of letters (AB or BA) and a statement about their sequential arrangement are presented to the subject. The subject's task is to determine whether the statement and letter pairs match or fail to match. For example, if a subject was presented with the statement "A IS FOLLOWED BY B" and the letter pair BA, he should respond FALSE. On the other hand, the subject should respond TRUE to the following statement and letter pair "A IS PRECEDED BY B"--BA. Responses are recorded by pressing one of two buttons on a keypad that are labeled TRUE and FALSE, respectively.

The test contains 32 unique sentence/letter pair stimuli that will be presented in the center of a CRT screen. This test can be performed with or without feedback.

BACKGROUND

This section will provide a brief overview of grammatical reasoning tasks. Four different types of procedures will be covered and compared.

Wason (1961) employed sentences that described whether a stated number was odd or even. For example, "seventy-six is an even number" (true affirmative) or "seventy-six is not an odd number" (true negative). There were

24 sentences that combined affirmative/negative and true/false. Wason found that negative statements were verified more slowly than positives. This finding was interpreted to mean that negative statements required an "inversion" which led to the slower responses. For example, negative statements contain a supposition plus an assertion--the sentence "seventy-six is not an even number" supposes that seventy-six is even and then asserts this supposition is false. Thus, subjects would interpret "not even" as "odd."

Research by Slobin (1966) has also illustrated that subjects can verify positive sentences more rapidly than negative sentences. Slobin employed pictures (e.g., a cat chasing a dog, a girl watering a flower, a man eating watermelon, etc.) instead of numerical quantities. In this experiment the subjects listened to a sentence and then viewed a picture. The subject was to decide if the sentence was true or false with regard to the picture.

Clark and Chase (e.g., Chase and Clark, 1972; Clark and Chase, 1972; Clark and Chase, 1974) have extensively studied the cognitive processes under-lying the comparison of pictorial information against sentences. In their experiments, subjects are shown a picture (e.g., $^+_*$ or $^*_+$) which matches or fails to match the meaning of a sentence. For example, ($^+_*$) followed by "the star is not above the plus" should lead to the response "TRUE." Subjects were shown sentences that varied with respect to the following dimensions; (a) the word above or below, (b) true or false, and (c) positive or negative.

Clark and Chase found that negative sentences were responded to more slowly than positive sentences. The interpretation here was similar to Wason's. Negative sentences presumably involve an "inversion" (i.e., "not above" is interpreted as "below" which requires additional processing relative to positive sentences).

Baddeley (1968) developed the version of the test that is being implemented in the UTC-PAB. The test is based on the findings of Slobin (1966) and Wason (1961). Subsequent research by Baddeley and Hitch (e.g., Baddeley and Hitch, 1974; Hitch and Baddeley, 1976) has shown that subjects can

25

verify positive sentences more quickly than negative sentences. In addition, active sentences were verified more quickly than passive sentences. Slobin (1966) found similar results with respect to passive versus active sentences. Examples of the different grammatical forms of the verbal reasoning test used by Baddeley and Hitch are presented in Table 2.

TABLE 2. EXAMPLES OF DIFFERENT GRAMMATICAL
FORMS OF THE VERBAL REASONING TASK

Grammatical Form	Example
Active affirmative	A follows B
Active negative	A does not follow B
Passive affirmative	A is followed by B
Passive negative	A is not followed by B

Baddeley and Hitch (1974) and Hitch and Baddeley (1976) have shown that the grammatical reasoning test imposes relatively little demand on short term memory storage. For example, subjects were able to verify sentences just as quickly when they had to maintain and recall six letters (e.g., memory span for letters) as when no letters were presented for recall. However, performance on the reasoning task was degraded when subjects were required to articulate the digit series (the items to be recalled). This was interpreted to mean that the processing operations associated with short term memory storage rather than storage per se are critical in producing interference.

In summary, the UTC-PAB version of the grammatical reasoning task (traditional) is based on research involving sentence verification. This research has shown that positives are verified more quickly than negatives and passives more slowly than actives. These effects have been demonstrated with a variety of stimuli (e.g., complex and simple pictures, numbers, etc.) and procedures. Furthermore, this task appears to tap the processing component of working memory rather than its storage capacity.

26

RELIABILITY

Baddeley (1968) examined the test-retest reliability of a paper and pencil version of this test. There were 18 subjects that were tested twice on successive days. The average correlation between performance on the two days was +.80.

Carter, Kennedy, and Bittner (1981) have also examined the reliability of this test. Their study involved 36 subjects who were tested on 15 consecutive days (Saturdays and Sundays excluded). The test was a paper and pencil version similar to that employed by Baddeley (1968); however, the subjects were tested for 1 minute intervals instead of three. The response measure was the number of correct decisions over the 1 minute trials. The results of this study were as follows; (a) average performance increased linearly with practice, (b) the variances were stable over the 15 days of testing, (c) intertrial correlations tended to remain constant, especially after the fourth day of testing, and (d) the average intertrial correlation after day four was +.82. These results, along with those of Baddeley (1968), indicate that the grammatical reasoning test (e.g., the paper and pencil version) is a highly reliable test instrument.

The UTC-PAB version of this test differs from the above in that sentences will be presented one at a time on a CRT screen. This procedural variation will require that additional reliability studies be conducted. However, the above research (Carter et al., 1981) suggests that the grammatical reasoning task is robust to modifications in procedure. For example, a reliability coefficient of +.82 was obtained when trial duration was decreased to 1 minute. This is nearly equivalent to what was found by Baddeley (1968) with 3-minute trials. (It should be noted that decreasing the length of a test generally leads to a drop in reliability.)

VALIDITY

This test appears to measure "higher mental processes" associated with logical reasoning. For example, Baddeley (1968) reports a correlation of +.59 between performance on the grammatical reasoning test and the British

Army verbal intelligence test (N = 29). In addition, Carter, Kennedy, and Bittner (1981) found a correlation of +.44 between grammatical reasoning and the Wonderlic test of mental ability (N = 23). This evidence supports the notion that the grammatical reasoning test measures a subject's general logical reasoning ability.

This test also appears to measure the construct of working memory processing capability. As may be recalled, Baddeley and Hitch (1974) and Hitch and Baddeley (1976) found that a concurrent memory span task (recalling up to six letters) did not degrade performance on the grammatical reasoning task. However, when subjects were required to articulate the memory series performance, the reasoning task was adversely affected. It should be noted that articulation of the same word (e.g., "the-the-the..." or "one-two-three") did not affect performance on the reasoning task. This follows, since repetition of the same word should not require much in the way of short term memory processing.

In summary, the grammatical reasoning test appears to tap subject's logical reasoning ability. In addition, this test appears to measure working memory processing capacity rather than just its storage capacity (Baddeley and Hitch, 1974; Hitch and Baddeley, 1976).

SENSITIVITY

This test has been shown to be sensitive to the effects of sleep loss, environmental stress (e.g., performance under water), road pollutants, and diurnal variations. In addition, performance decrements in grammatical reasoning have been obtained in dual task experiments. Table 3 presents a list of various studies that have employed the grammatical reasoning task.

As can be seen in Table 3, the grammatical reasoning test appears to be sensitive to the effects of sleep loss and diurnal variations. However, one study (Pleban et al., 1985) did not report an effect of sleep loss on performance in the grammatical reasoning task. In this study the focus was

TABLE 3.   SAMPLE OF STUDIES UTILIZING THE GRAMMATICAL REASONING TASK

Reference	Factor Under Study	Reported Effect	N
Baddeley et al., 1968	Nitrogen Narcosis and Performance Under Water	Yes	18
Brown et al., 1968	Dual Task: Driving	Yes	24
Lewis et al., 1970	Traffic Pollution	Yes	15
Baddeley et al., 1975	Hypothermia (in divers)	No	14
Folkard, 1975	Diurnal Variations (time of day effects)	Yes	36
Poulton et al., 1978	Sleep Loss	Yes	14
Webb and Levy, 1984	Sleep Loss	Yes	6
Angus and Heslegrave, 1985	Sleep Loss	Yes	12
Englund et al., 1985	Diurnal Variations (time of day effects)	Yes	22
Pleban et al., 1985	Sleep Loss and Physical Fitness	No	16

on the correlation between changes in performance on cognitive tests (e.g., grammatical reasoning, map-plotting test, and encoding-decoding test) as a function of sleep loss and measures of physical fitness (e.g., chin-ups, push-ups, sit-ups, two-mile run, and pulse rate). The study reports that there was not a statistically reliable relationship between physical fitness and performance decrements on the grammatical reasoning test as a function of sleep loss. However, performance on the grammatical reasoning test may have been sensitive to the effects of sleep loss per se, but the manner in which the results are reported makes this determination impossible.

The grammatical reasoning test has also been shown to be sensitive to the effects of environmental stressors (e.g., performance under water), and

toxic substances (e.g., traffic pollution). However, a study by Baddeley et al. (1975) showed that highly motivated subjects were unimpaired on the grammatical reasoning test despite a marked drop in core temperature (performance on a vigilance task was also unimpaired).

Finally, the grammatical reasoning task has been shown to affect performance on a driving task in a dual task paradigm (Brown et al., 1968). In this study subjects responded "true" or "false" via a car phone to auditorially presented sentences (the researchers were interested in determining the effects of communicating on a car phone with driving performance). The grammatical reasoning task mainly impaired judgements of "impossible" gaps (gaps which were smaller than the car). However, the control skills employed in steering through "possible" gaps (gaps that were larger than the car) were not readily degraded, although speed of driving was significantly reduced.

The above indicates that the grammatical reasoning test is highly sensitive to the effects of environmental stressors, toxic substances, and the demands imposed by a demanding concurrent task. However, the research by Baddeley et al. (1985) points out the importance of motivational factors in the evaluation of performance under stress.

TECHNICAL DESCRIPTION

The stimulus items differ on five binary dimensions, yielding 32 unique combinations. These dimensions are: (1) positive or negative statement, (2) active or passive voice, (3) follow or precede verb root, (4) AB or BA letter pair, and (5) A...B or B...A order within the statement. A sixth dimension redundantly determined by the above is whether the statement-pair relationship is true or false. The eight base sentences described in terms of the above dimenions are presented on Table 4.

Stimulus items occupy the center five lines of the display. The first line displays the sentence. The second is blank. The third contains a solid nonblinking cursor to serve as a reference point, prompt, and feedback symbol. The fourth is blank. The fifth contains the letter pair "AB" or

30

TABLE 4. EIGHT BASE SENTENCES

Sentence	Letter Pair	Dimensions		
___ Follows ___	___	POS	ACT	FOL
___ Precedes ___	___	POS	ACT	PRE
___ Is Followed By ___	___	POS	PAS	FOL
___ Is Preceded By ___	___	POS	PAS	PRE
___ Does Not Follow ___	___	NEG	ACT	FUL
___ Does Not Precede ___	___	NEG	ACT	PRE
___ Is Not Followed By ___	___	NEG	PAS	FOL
___ Is Not Preceded By ___	___	NEG	PAS	PRE

"BA." All lines are centered. All characters are upper case. Display colors are white characters on a light blue background with a dark blue border.

Valid responses are presses of the true or false buttons. Invalid responses are recorded as "extras" but have no other effect. If no valid response occurs for 15 seconds a beep is sounded, the screen is blanked for 1000 msec and the next trial continues.

Trial Specifications

Each trial will involve the following steps; (a) a sentence/letter pair stimulus is presented until a valid response (TRUE or FALSE key is pressed) is entered or 15 seconds elapse; (b) the screen is cleared; (c) the word CORRECT OR INCORRECT is displayed in the center of the CRT for 1000 msec or if no feedback option is selected, the screen remains blank for 500 msec; (d) the screen is cleared if the feedback option was selected. The above process is repeated for each of the 32 stimuli in this test.

DATA SPECIFICATIONS

Each trial records a stimulus code, a response code, and a reaction time value. The stimulus code identifies the item in terms of the six

dimensions mentioned above: (1) positive or negative statement, (2) active or passive voice, (3) follow or precede verb root, (4) AB or BA letter pair, (5) A...B or B...A order within the sentences, and (6) whether the sentence letter pair was TRUE or FALSE. The response code identifies whether the subject pressed the TRUE button or the FALSE button, and whether the response was correct, incorrect, or terminated by the deadline. The reaction time value is the time from the stimulus presentation to the occurrence of the response, or is set equal to the deadline value.

Summary data are: (1) total elapsed time (task duration in seconds), (2) number of trials completed, (3) number and percent correct, (4) number of extras, (5) number of deadline occurrences, and (6) reaction time means and standard deviations for total responses, correct responses, and incorrect responses (not counting deadlines or extras). The review of the literature suggests that average reaction time for correct responses and number of errors can serve as the major dependent measures.

TRAINING REQUIREMENTS

Following the instructions the subjects should receive a minimum of 10 practice trials. The practice trials should provide feedback with respect to speed and accuracy for each trial. In addition, the feedback should remain visible until the subject presses a key to start the next trial sequence. Providing feedback after each trial and placing the practice trials under subject control will increase the likelihood of subjects understanding and following directions during the experimental trials. In addition, subject paced trials will allow the experimenter to carefully monitor performance during practice and to answer questions that subjects may have regarding the nature of the task.

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

In this task you will be presented with a series of statements about the relationship between two letters. Each statement will be followed by the letter pair AB or BA. Your task is to determine whether the statement correctly describes the order of the letters within the pair.

For example, if you were to see the statement "A is followed by B" with the letter pair AB, you should respond "true" by pressing the button marked "true." On the other hand, if you were to see the statement "A is not preceded by B" with the letter pair BA, you should respond "false" by pressing the button labeled "false."

For this task it is important that you make your decisions as quickly and as accurately as you can. If you take more than 15 seconds to make a response, a tone will be sounded and the computer will go on to the next trial.

You will now be presented with a series of 10 practice trials. If you are not sure of the answer, ask for clarification. Many people have difficulty at first with some of the relationships.

33

Section 4
GRAMMATICAL REASONING (SYMBOLIC) (UTC-PAB TEST NO. 3)
(LOGICAL REASONING)

PURPOSE

The purpose of this task is to tap resources dedicated to general reasoning
ability. The symbolic grammatical reasoning task is a type of sentence
verification task that taps the processing capacity of working memory.
This task is known to be sensitive to variable information processing
demands and is probably sensitive to environmental stress, pollutants, and
sleep loss.

DESCRIPTION

The symbolic grammatical reasoning task is designed to impose variable
demands on resources required for the manipulation and comparison of gram-
matical information. The task is derived from Baddeley's (1968) Grammati-
cal Reasoning Task. The stimuli consist of sentences of varying syntactic
structure accompanied by sets of two or three simultaneously presented sym-
bols (e.g., *, @, and #). The sentences must be analyzed to determine
whether they correctly describe the ordering of the characters in the sym-
bol set. Task demand is determined by the amount and complexity of gram-
matical analysis. Three different levels of task demand are imposed by the
following task conditions: (1) single-sentence items of variable syntactic
construction describing the order of pairs of symbols (i.e., all possible
stimuli from the Baddeley version, substituting symbols for the letters)--
low demand; (2) items composed of two sentences worded actively and posi-
tively, describing the relative positions of three symbols--moderate
demand; and (3) two-sentence items worded either actively/negatively or
passively/negatively and describing three symbols--high demand. Figure 3
shows mean reaction times and subjective difficulty ratings associated with
these conditions (Shingledecker, 1984).

GRAMMATICAL REASONING DATA
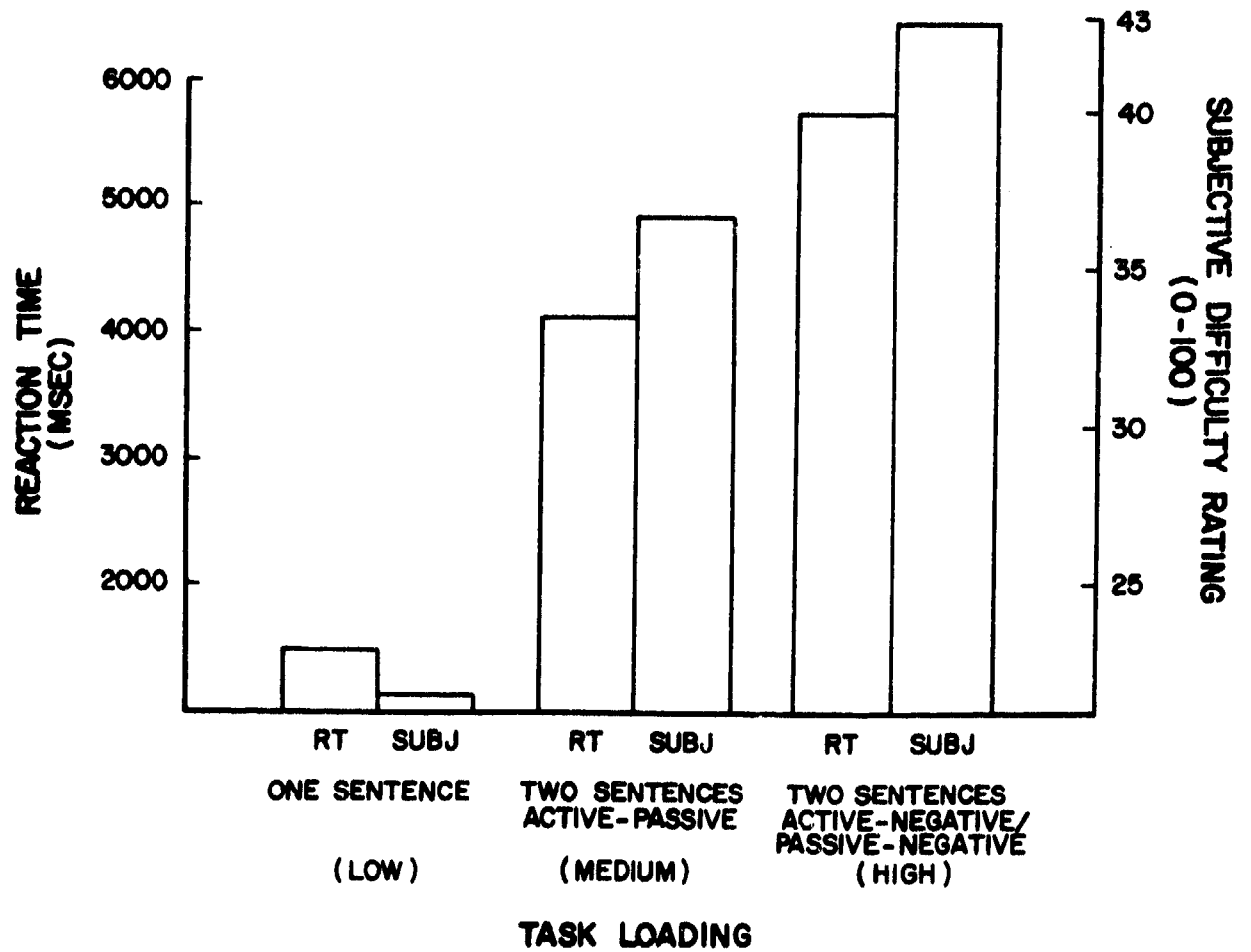
Figure 3.   Grammatical Reasoning Data

BACKGROUND

This section will provide a brief overview of grammatical reasoning tasks
found in the literature. Five different types of procedures will be
covered and compared.

Wason (1961) employed sentences that described whether a stated number was
odd or even. For example, "seventy-six is an even number" (true affirm-
ative) or "seventy-six is not an odd number" (true negative). There were
24 sentences that combined affirmative/negative and true/false. Wason
found that negative statements were verified more slowly than positives.
This finding was interpreted to mean that negative statements required an
"inversion" which led to the slower responses. For example, negative
statements contain a supposition plus an assertion--the sentence "seventy-
six is not an even number" supposes that seventy-six is even and then
asserts this supposition is false. Thus, subjects would interpret "not
even" as "odd."

Research by Slobin (1966) has also illustrated that subjects can verify
positive sentences more rapidly than negative sentences. Slobin employed
pictures (e.g., a cat chasing a dog, a girl watering a flower, a man eating
watermelon, etc.) instead of numerical quantities. In this experiment the
subjects listened to a sentence and then viewed a picture. The subject was
to decide if the sentence was true or false with regard to the picture.

Clark and Chase (e.g., Chase and Clark, 1972; Clark and Chase, 1972; Clark
and Chase, 1974) have extensively studied the cognitive processes underly-
ing the comparison of pictorial information against sentences. In their
experiments, subjects are shown a picture (e.g., $\overset{*}{+}$ or $\overset{*}{\ddagger}$)which matches or
fails to match the meaning of a sentence. For example, ($\overset{*}{+}$) followed by
"the star is not above the plus" should lead to the response "TRUE." Sub-
jects were shown sentences that varied with respect to the following dimen-
sions; (a) the word above or below, (b) true or false, and (c) positive or
negative.

Clark and Chase found that negative sentences were responded to more slowly than positive sentences. The interpretation here was similar to Wason's. Negative sentences involve an "inversion" (i.e., "not above" is interpreted as "below" which required additional processing relative to positive sentences).

Baddeley (1968) developed the traditional version of the task (UTC-PAB Test No. 2) which was based on the findings of Slobin (1966) and Wason (1961). Subsequent research by Baddeley and Hitch (e.g., Baddeley and Hitch, 1974; Hitch and Baddeley, 1976) has shown that subjects can verify positive sentences more quickly than negative sentences. In addition, active sentences were verified more quickly than passive sentences (Slobin found similar results with respect to passive versus active sentences).

Baddeley and Hitch (1974) and Hitch and Baddeley (1976) have shown that the grammatical reasoning test imposes relatively little demand on short term memory store. For example, subjects were able to verify sentences just as quicky when they had to maintain and recall six letters (e.g., memory span for letters) as when no letters were presented for recall. However, performance on the reasoning task was degraded when subjects were required to articulate the digit series (the items to be recalled). This was interpreted to mean that the processing operations associated with short term memory storage rather than storage per se are critical in producing interference.

The version of the traditional grammatical reasoning task as it appears in the UTC-PAB (Test No. 2) is based on research involving sentence verification. This research has shown that positives are verified more quickly than negatives and passives more slowly than actives. These effects have been demonstrated with a variety of stimuli (e.g., complex and simple pictures, numbers, etc.) and procedures. Furthermore, this task appears to tap the processing component of working memory rather than its storage capacity.

The symbolic version of the grammatical reasoning task was originally developed by Shingledecker (1984). This version of the task represents an

attempt to combine elements of Baddelely's (1968) often cited traditional task, as per UTC-PAB Test No. 2, with elements of the Clark and Chase (Chase and Clark, 1972; Clark and Chase, 1972; Clark and Chase, 1974) studies to produce a paradigm that is potentially of greater diagnosticity, for some purposes, than either of its antecedent paradigms. The underlying rationale of this integration lies with a concern for maximal construct validity, which is very important in performance assessment research. The construct of interest for this task is logical reasoning. In other words, it is imperative that the subjects utilize the informtion contained within the stimulus sentences to make their logical determinations. Only then can the various task loadings be said to differentially affect central processing resources dedicated to logical reasoning ability. It occurred to Shingledecker (1984) that the use of letter pairs as the target set (Baddeley, 1968) may, at times, lessen the degree to which a subject must depend upon the logical structure of the sentence(s). For example, the letters A and B bear with them a natural alphabetic order, and a subject could simply encode the target set as "right" (i.e., AB) or "wrong" (i.e., BA) instead of "A precedes B," etc. It would seem then that a portion of the logical reasoning process can be bypassed by developing working memory chunking strategies which center around the target letters themselves. The employment of the less verbally meaningful symbols *, #, and @ (Chase and Clark, 1972; Clark and Chase, 1972; Clark and Chase, 1974) in this paradigm, instead of letters, should alleviate this problem.

The question then becomes "which grammatical reasoning paradigm is the one to use, UTC-PAB Test No. 2 or No. 3?" The answer is that this decision involves some tradeoffs that have been implied previously. The traditional version (UTC-PAB Test No. 2) may be characterized by the potential construct validity confound cited above. However, this has not been steadfastly proven and a considerable amount of research has been conducted with this paradigm. As is mentioned elsewhere, the literature indicates that a high degree of reliability, validity, and sensitivity are associated with the traditional version of this test, and these dimensions are very important in performance assessment research.

If a given testing situation is such that construct validity is paramount, UTC-PAB Test No. 3, symbolic grammatical reasoning, may be viewed as a better alternative to avoid the potential problems which may beset the use of letter pairs (as noted by Shingledecker, 1984). The disadvantage here is that no reliability, validity, or sensitivity data can be specifically related to this paradigm, though there is reason to suspect that the task would be characterized by a sufficient degree of all three dimensions (see sections on realiability, validity, and sensitivity). In summary, each version seems to have its relative merits, although additional research specifically investigating the issues discussed here is required before any conclusions can be drawn.

RELIABILITY

Baddeley (1968) examined the test-retest reliability of his traditional paper and pencil version of this test. Eighteen subjects were tested twice on succssive days, yielding an average correlation between performance on the two days of +.80.

Carter, Kennedy, and Bittner (1981) have also examined the reliability of the grammatical reasoning test. Thirty-six subjects were tested on 15 con-secutive workdays. The test employed was a paper and pencil version of the traditional grammatical reasoning pagadigm, similar to that employed by Baddeley (1968); however, the subjects were tested for 1 minute invervals instead of three. The response measure incorporated into the analyses was the number of correct determinations per each 1 minute trial.

Carter et al. (1981) found that: (a) average performance increased lin-early with practice, (b) the variances were stable over the 15 days of testing, (c) intertrial correlations tended to remain constant, especially after the fourth day of testing, and (d) the average intertrial correlation after day 4 was +.82. These results, along with those of Baddeley (1968), indicate that the paper and pencil version of the traditional grammatical reasoning task (UTC-PAB Test No. 2) is a very reliable test instrument and, thus, suggest that the symbolic grammatical reasoning paradigm should be as well.

39

This task differs from those found reliable by Baddeley (1968) and Carter et al. (1981) in that: (1) sentences will be presented one at a time on a CRT screen, and (2) the symbols #, @, and # will be used instead of the letters A and B. These procedural variations will require that additional reliability studies be conducted. However, the aforementioned research (Baddeley, 1968; Carter et al., 1981) implies that this task is robust to procedural variation, as the reliability coefficient (+.82) obtained with one-minute trials (Carter et al., 1981) is nearly equivalent to that obtained with 3-minute trials (Baddeley, 1968; it should be noted that decreasing the duration of a test generally leads to decreased reliability).

VALIDITY

This test likely taps into the "higher mental processes" associated with logical reasoning. Baddeley (1968) reports a correlation of +.59 between performance on the paper and pencil version of the traditional grammatical reasoning task (UTC-PAB Test No. 2) and the British Army Verbal Intelligence Test (N = 29). Using a similar version of the task Carter, Kennedy, and Bittner (1981) obtained a correlation of + .44 between the grammatical reasoning test and the Wonderlic Test of Mental Ability (N = 23). These findings support the notion that this grammatical reasoning paradigm measures a subject's general logical reasoning ability.

The traditional grammatical reasoning test also appears to measure the construct of working memory processing ability. Baddeley and Hitch (1974) found that a concurrent memory span task (recalling up to 6 letters) did not degrade grammatical reasoning performance. However, when subjects were required to articulate the memory series, reasoning performance was adversely affected. It should be noted that articulation of the same word (e.g., the-the-the...) or a redundant series (e.g. one-two-three) did not affect reasoning performance. These results were interpreted to mean that the processing operations associated with short term memory storage, rather than storage per se, are critical in producing interference on the traditional grammatical reasoning task.

In summary, the traditional grammatical reasoning test appears to tap processing resources dedicated to logical reasoning ability and working memory processing capacity rather than just storage capacity (Baddeley and Hitch, 1974). Though such a study has yet to be conducted utilizing the symbolic grammatical reasoning task, these investigations involving traditional grammatical reasoning can be interpreted to suggest that the symbolic test would be characterized by a correspondingly significant degree of construct validity.

SENSITIVITY

The sensitivity of the symbolic version of the grammatical reasoning paradigm has not yet been conclusively investigated. Such investigations, though, would be very informative for reasons discussed previously in the Background section. Due to the lack of specifically pertinent research, the sensitivity of the traditional grammatical reasoning paradigm will be discussed here, for it is likely that the employment of the symbolic version would produce similar findings, though the actual employment of the symbolic test is required to truly assess its sensitivity.

The traditional grammatical reasoning paradigm (UTC-PAB Test No. 2) has been shown to be sensitive to the effects of sleep loss, environmental stressors (e.g., performance under water), road pollutants, and diurnal variations. Performance decrements in grammatical reasoning have been obtained when a dual task paradigm is employed. Table 2 presented a list of various studies that have employed the traditional grammatical reasoning task.

As can be seen in Table 5, the traditional grammatical reasoning test appears to be highly sensitive to the effects of sleep loss and diurnal variations. However, one study (Pleban et al., 1985) did not report an effect of sleep loss on performance in the grammatical reasoning task. In this study the focus was on the correlation between changes in performance on cognitive tests (e.g., grammatical reasoning, map-plotting test, and encoding-decoding test) as a function of sleep loss and measures of physical fitness (e.g., chin-ups, push-ups, sit-ups, two-mile run, and pulse

41

rate). The study reports that there was not a statistically reliable rela-
tionship between physical fitness and performance decrements on the gram-
matical reasoning test as a function of sleep loss. However, performance
on the grammatical reasoning test may have been sensitive to the effects of
sleep loss, but the manner in which the results are reported make this
determination impossible.

TABLE 5.   STUDIES UTILIZING THE GRAMMATICAL REASONING TASK

References	Factor Under Study	Reported Effect	N
Baddeley et al., 1968	Nitrogen Narcosis and Performance Under Water	Yes	18
Brown et al., 1968	Dual Task: Driving	Yes	24
Lewis et al., 1970	Traffic Pollution	Yes	15
Baddeley et al., 1975	Hypothermia (in divers)	No	14
Folkard, 1975	Diurnal Variations (time of day effects)	Yes	36
Poulton et al., 1978	Sleep Loss	Yes	14
Webb and Levy, 1984	Sleep Loss	Yes	6
Angus and Heslegrave, 1985	Sleep Loss	Yes	12
Englund et al., 1985	Diurnal Variations (time of day effects)	Yes	22
Pleban et al., 1985	Sleep Loss and Physical Fitness	No	16

The grammatical reasoning test has also been shown to be sensitive to the
effects of environmental stressor (e.g., performance under water) and toxic
substances (e.g., traffic pollution). However, a study by Baddeley et al.
(1975) showed that highly motivated subjects were unimpaired on the gram-
matical reasoning test despite a marked drop in core temperature. (Per-
formance on a vigilance task was also unimpaired.)

Finally, the traditional grammatical reasoning task has been shown to
affect performance on a driving task in a dual task paradigm (Brown et al.,
1968). In this study subjects responded "true" or "false" via a car phone
to auditorially presented sentences (the researchers were interested in
determining the effects of communicating on a car phone with driving per-
formance). The grammatical reasoning task mainly impaired judgements of
"impossible" gaps (gaps which are smaller than the car). However, the con-
trol skills employed in steering through "possible" gaps (gaps that were
larger than the car) were not readily degraded, although speed of driving
was significantly reduced.

The above indicates that the traditional grammatical reasoning test (UTC-
PAB Test No. 2) is highly sensitive to the effects of environmental stress-
ors, toxic substances, and the demands imposed by a demanding concurrent
task and suggests that the findings may be similar if the symbolic version
of the test had been employed. However, the research by Baddeley et al.
(1985) showed that the performance of highly motivated subjects was not
affected by extreme cold. This result points out the importance of motiva-
tional factors in the evaluation of performance under stress.

TECHNICAL DESCRIPTION

The stimulus population for single sentence (low demand) problems is com-
prised of all 32 possible combinations of the following five binary fac-
tors: (1) active versus passive wording of sentences; (2) positive versus
negative wording; (3) keyword "follows" versus "precedes"; (4) order of the
two symbols in the sentence; and (5) order of symbols in the symbol set.
All 32 possible one-sentence test items are shown in Table 6. For one sen-
tence item, the subject's task is to decide whether the symbol set is
ordered as the sentence indicates and respond either positively or
negatively.

In the task conditions using two sentences (medium and high task demand),
the subject is required to determine whether the sentences match in their
assessment of the symbol set. If both sentences correctly describe the
ordering of the three symbols, or if neither is correct, the subject should

43

TABLE 6. GRAMMATICAL REASONING ITEMS FOR THE LOW DEMAND CONDITION

CTS GRAMMATICAL REASONING TASK DATA FOR LEVEL 1 (ONE SENTENCE)

NUMBER	SENTENCE	SYMBOL	ANSWER
1	@ PRECEDES *	@*	MATCH
2	@ FOLLOWS *	@*	NONMATCH
3	@ IS PRECEDED BY *	@*	NONMATCH
4	@ IS FOLLOWED BY *	@*	MATCH
5	@ DOES NOT PRECEDE *	@*	NONMATCH
6	@ DOES NOT FOLLOW *	@*	MATCH
7	@ IS NOT PRECEDED BY *	@*	MATCH
8	@ IS NOT FOLLOWED BY *	@*	NONMATCH
9	@ PRECEDES *	*@	NONMATCH
10	@ FOLLOWS *	*@	MATCH
11	@ IS PRECEDED BY *	*@	MATCH
12	@ IS FOLLOWED BY *	*@	NONMATCH
13	@ DOES NOT PRECEDE *	*@	MATCH
14	@ DOES NOT FOLLOW *	*@	NONMATCH
15	@ IS NOT PRECEDED BY *	*@	NONMATCH
16	@ IS NOT FOLLOWED BY *	*@	MATCH
17	* PRECEDES @	@*	NONMATCH
18	* FOLLOWS @	@*	MATCH
19	* IS PRECEDED BY @	@*	MATCH
20	* IS FOLLOWED BY @	@*	NONMATCH
21	* DOES NOT PRECEDE @	@*	MATCH
22	* DOES NOT FOLLOW @	@*	NONMATCH
23	* IS NOT PRECEDED BY @	@*	NONMATCH
24	* IS NOT FOLLOWED BY @	@*	MATCH
25	* PRECEDES @	*@	MATCH
26	* FOLLOWS @	*@	NONMATCH
27	* IS PRECEDED BY @	*@	NONMATCH
28	* IS FOLLOWED BY @	*@	MATCH
29	* DOES NOT PRECEDE @	*@	NONMATCH
30	* DOES NOT FOLLOW @	*@	MATCH
31	* IS NOT PRECEDED BY @	*@	MATCH
32	* IS NOT FOLLOWED BY @	*@	NONMATCH

respond positively. If one sentence is correct but the other is not, a negative response is required. Sentences always describe adjacent symbol pairs and are of the same grammatical form (i.e., an active/negative sentence is never paired with a passive/negative sentence). To help balance all conditions, sets of 32 grammatical problems are randomly chosen from the larger stimulus populations associated with two-sentence items. Two restrictions are imposed on this selection process: (1) when correctly solved, half of the two-sentence problems must necessitate a positive

44

response, and (2) combinations of sentence solutions (e.g., sentence one "true," sentence two "true,"; sentence one "true," sentence two "false," etc.) are to occur equally often. Equal numbers of active/negative and passive/negative items appear in the high demand condition. Two-sentence test items for the moderate and high task demand conditions are shown in Tables 7 and 8, respectively. During experimental trials, the computer randomly selects test items from the appropriate list for presentation. Also during testing, response deadlines vary with task loading (as will resulting RTs; Shingledecker, 1984). The deadline for the low demand condition (simple one-sentence items) is 2.5 seconds; for the moderate demand condition (two sentences, active/positive wording) 6.5 seconds; and for the high demand condition (two sentences, active/negative or passive/negative wording) 7.5 seconds. Binary responses are entered manually on two appropriately labeled keys on a four button keypad.

DATA SPECIFICATIONS

Recorded for each trial are a stimulus code, a response code, and a reaction time value. The stimulus code identifies the item it terms of six possible stimulus dimensions: (1) positive or negative statement, (2) active or passive voice, (3) "follow" or "precede" verb root, (4) symbol set (e.g., *@, @*, *@#, etc.), (5) specific order of symbols within the sentences, and (6) whether a sentence is TRUE or FALSE. The response code identifies whether the subject pressed the TRUE key or the FALSE key, and whether this response is correct, incorrect, or terminated by the given response deadline. Reaction time is measured from the onset of stimulus presentation to the occurrence of the response, or is set equal to the deadline value if the reaction time is in excess of the deadline.

Summary data are: (1) total elapsed time (task duration in seconds), (2) number of trials completed, (3) number and percent correct, (4) number of extras, (5) number of deadline occurrences, and (6) reaction time means and standard deviations for total responses, correct responses only, and incorrect responses only (excluding deadlines and extras as well). Average reaction time for correct responses and number of errors usually serve as the major dependent measures in the grammatical reasoning paradigm.

TABLE 7. GRAMMATICAL REASONING ITEMS FOR THE MODERATE DEMAND CONDITION

DATA FOR LEVEL 2 (TWO SENTENCE-AP)				
NUMBER	SENTENCE 1	SENTENCE 2	SYMBOL	ANSWER
1	@ PRECEDES *	@ FOLLOWS #	*@#	MATCH
2	# FOLLOWS *	@ PRECEDES *	#*@	MATCH
3	* PRECEDES #	@ FOLLOWS #	@#*	MATCH
4	# PRECEDES @	* PRECEDES #	*#@	MATCH
5	# PRECEDES @	@ PRECEDES *	*@#	MATCH
6	@ PRECEDES #	* FOLLOWS #	@#*	MATCH
7	@ PRECEDES #	* PRECEDES @	*@#	MATCH
8	@ FOLLOWS #	@ PRECEDES *	#@*	MATCH
9	# FOLLOWS @	@ FOLLOWS *	*@#	MATCH
10	* FOLLOWS @	* PRECEDES #	@*#	MATCH
11	* FOLLOWS @	@ FOLLOWS #	*@#	MATCH
12	* FOLLOWS @	# FOLLOWS *	@*#	MATCH
13	# PRECEDES *	* PRECEDES @	@*#	MATCH
14	* FOLLOWS #	@ PRECEDES #	*#@	MATCH
15	@ PRECEDES *	* PRECEDES #	#*@	MATCH
16	# PRECEDES *	@ FOLLOWS *	@*#	MATCH
17	* FOLLOWS @	@ PRECEDES #	#@*	NONMATCH
18	* FOLLOWS #	@ PRECEDES *	@*#	NONMATCH
19	* PRECEDES @	# FOLLOWS *	@*#	NONMATCH
20	@ FOLLOWS #	* PRECEDES @	#@*	NONMATCH
21	# FOLLOWS @	* PRECEDES #	*#@	NONMATCH
22	# PRECEDES @	@ FOLLOWS *	#@*	NONMATCH
23	# FOLLOWS @	@ PRECEDES *	*@#	NONMATCH
24	* FOLLOWS @	# PRECEDES *	#*@	NONMATCH
25	@ PRECEDES *	# FOLLOWS @	#@*	NONMATCH
26	* PRECEDES #	@ FOLLOWS *	#*@	NONMATCH
27	# PRECEDES @	* FOLLOWS #	@#*	NONMATCH
28	@ FOLLOWS #	# PRECEDES *	@#*	NONMATCH
29	# FOLLOWS *	@ PRECEDES #	*#@	NONMATCH
30	* PRECEDES #	# FOLLOWS @	*#@	NONMATCH
31	@ FOLLOWS *	# PRECEDES @	*@#	NONMATCH
32	@ FOLLOWS *	* PRECEDES #	@*#	NONMATCH

Note: If both sentences are true or both are false, the correct answer is MATCH. On the other hand, if one sentence is true and the other false, the correct answer is NONMATCH.

TABLE 8. GRAMMATICAL REASONING ITEMS FOR THE HIGH DEMAND CONDITION

DATA FOR LEVEL 3 (TWO SENTENCE-AN/PN)

NUMBER	SENTENCE 1	SENTENCE 2	SYMBOL	ANSWER
1	# DOES NOT PRECEDE *	# DOES NOT FOLLOW @	*#@	MATCH
2	* DOES NOT FOLLOW @	@ DOES NOT FOLLOW #	*@#	MATCH
3	* DOES NOT PRECEDE #	* DOES NOT FOLLOW @	#*@	MATCH
4	@ DOES NOT PRECEDE *	# DOES NOT FOLLOW *	#*@	MATCH
5	@ DOES NOT PRECEDE #	@ DOES NOT FOLLOW *	*@#	MATCH
6	@ DOES NOT FOLLOW #	* DOES NOT PRECEDE #	*#@	MATCH
7	@ DOES NOT FOLLOW #	# DOES NOT FOLLOW *	*#@	MATCH
8	* DOES NOT PRECEDE @	# DOES NOT PRECEDE *	#*@	MATCH
9	# DOES NOT PRECEDE *	@ DOES NOT FOLLOW #	@#*	NONMATCH
10	# DOES NOT FOLLOW @	@ DOES NOT PRECEDE *	#@*	NONMATCH
11	# DOES NOT FOLLOW *	@ DOES NOT PRECEDE #	@#*	NONMATCH
12	# DOES NOT PRECEDE @	@ DOES NOT FOLLOW *	#@*	NONMATCH
13	* DOES NOT FOLLOW #	# DOES NOT PRECEDE @	*#@	NONMATCH
14	# DOES NOT PRECEDE *	* DOES NOT FOLLOW @	@*#	NONMATCH
15	* DOES NOT PRECEDE @	@ DOES NOT FOLLOW #	*@#	NONMATCH
16	* DOES NOT FOLLOW @	@ DOES NOT PRECEDE #	#@*	NONMATCH
17	* IS NOT PRECEDED BY #	# IS NOT PRECEDED BY @	@#*	MATCH
18	* IS NOT PRECEDED BY @	# IS NOT FOLLOWED BY @	#@*	MATCH
19	@ IS NOT PRECEDED BY *	# IS NOT FOLLOWED BY *	@*#	MATCH
20	@ IS NOT PRECEDED BY #	@ IS NOT FOLLOWED BY *	*@#	MATCH
21	@ IS NOT FOLLOWED BY #	@ IS NOT PRECEDED BY *	#@*	MATCH
22	@ IS NOT FOLLOWED BY #	* IS NOT PRECEDED BY #	@#*	MATCH
23	# IS NOT PRECEDED BY *	@ IS NOT PRECEDED BY #	@#*	MATCH
24	# IS NOT PRECEDED BY @	* IS NOT FOLLOWED BY @	*@#	MATCH
25	* IS NOT FOLLOWED BY @	@ IS NOT PRECEDED BY #	#@*	NONMATCH
26	* IS NOT FOLLOWED BY #	@ IS NOT PRECEDED BY *	#*@	NONMATCH
27	@ IS NOT FOLLOWED BY *	# IS NOT PRECEDED BY @	#@*	NONMATCH
28	* IS NOT FOLLOWED BY @	# IS NOT PRECEDED BY *	#*@	NONMATCH
29	@ IS NOT PRECEDED BY #	# IS NOT FOLLOWED BY *	@#*	NONMATCH
30	# IS NOT FOLLOWED BY @	@ IS NOT PRECEDED BY *	#@*	NONMATCH
31	* IS NOT PRECEDED BY @	# IS NOT FOLLOWED BY *	#*@	NONMATCH
32	* IS NOT PRECEDED BY #	# IS NOT FOLLOWED BY @	@#*	NONMATCH

Note: If both sentences are true or both are false, the correct answer is MATCH. On the other hand, if one sentence is true and the other false, the correct answer is NONMATCH.

TRAINING REQUIREMENTS

Subjects are presented with the instructions. Two 36-minute training
sessions composed of four 3-minute trials at each level of task difficulty
are suggested.

During training, presentation of grammatical problems is subject-paced with
a 15-second deadline for all three demand levels. If the subject does not
respond within 15 seconds of the onset of the stimulus, the display is
cleared and a new item is presented. Subjects should receive performance
feedback throughout the training trials to maintain acceptance performance
levels.

To summarize, the training phase for this test should consist of the
following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that
    the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require
    additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over
    several sessions on this test, one may omit the practice trials after
    the first session.

INSTRUCTIONS TO SUBJECTS

You will be presented with sentences that vary in their structural complex-
ity. Each sentence contains two symbols, and either correctly or incor-
rectly describes the order of the symbols as they appear adjacent to the
sentence. Your task is to determine as quickly and accurately as possible
whether the sentences correctly describe the order of the symbols, and

then, based on this determination, press the "yes" or "no" button on the keypad.

There are three categories of grammatical reasoning problems. The first category is composed of single-sentence problems which describe the order of two symbols. In the single-sentence condition, you are to describe whether the sentence accurately reflects the order of the two symbols. In the example problem:

<div align="center">

* IS PRECEDED BY @   @*

</div>

The * is, in fact, preceded by the @, so the correct response would be "yes." The structure of the sentences in the single-sentence condition is variable. That is, sometimes the sentence will be worded simply and sometimes not..

The second category of task problems is composed of pairs of sentences which describe the ordering of three symbols. The sentence wording at this level of the task is always simple. Your task is to determine whether both sentences are correct or incorrect, or whether one sentence is correct while the other is incorrect. If one sentence is correct and the other not, you should respond "nonmatch" (or "no"). If both are either correct or incorrect, you should respond "match" (or "yes"). For example:

<div align="center">

PRECEDES @ ⎫
          ⎬   #@*
* FOLLOWS @ ⎭

</div>

The # does precede the @, so the first sentence is correct and the * does follow the @, so the second sentence is also correct. Since both sentences are correct (rather than one correct and one incorrect) the sentence answers match, and the appropriate response is "same."

In the third task category, two sentences again describe the order of three symbols, but the sentences are worded in a more complicated fashion. As in

<div align="center">

49

</div>

the other two-sentence condition, the task is to compare the correctness of the sentences. For example:

```
        * IS NOT PRECEDED BY @
                                        @*#
        # IS PRECEDED BY *
```

In this case the * is preceded by the @, so the first sentence is incorrect and the # is preceded by the *, so the second sentence is correct. Since one sentence is correct but the other not, the correct response would be "different."

You should try to respond as quickly and accurately as you can to each problem. If you find yourself making repeated errors because you are not taking enough time for your decision, slow down. However, do not take more time than is necessary to make the appropriate decision and response. You will start the experimental session by pressing a key on the response keypad. The trials will last 3 minutes each. At the end of 3 minutes the task will stop by itself and the screen will go blank.

Section 5
TWO-COLUMN ADDITION (UTC-PAB TEST NO. 4)
(NUMBER FACILITY)

PURPOSE

The purpose of this subject-paced, mental arithmetic test is to measure the subject's ability to sum simple addition problems. The test is diagnostic of the speed and accuracy with which subjects retrieve arithmetic information (e.g., math facts) and utilize procedural knowledge (e.g., well learned procedures for adding columns of digits). In addition, short term storage of carry and intermediate result information is required.

DESCRIPTION

During this arithmetic test, a set of 45 trials is presented to the subject. Each trial consists of three 2-digit numbers being presented on a CRT screen simultaneously in a column format. The subject is required to sum as rapidly as possible and enter his/her response via a keyboard. Responses must be entered beginning with the left hand digit first (usually the hundreds and tens digit). The column of digits displayed on the CRT screen will disappear with the first valid key entry; thus, subjects must know the entire answer prior to entering a response. A trial ends when the return key is pressed or when a deadline period of 15 seconds has passed. Subjects will receive speed/accuracy feedback during the training trials; however, no feedback will be provided during the experimental trials.

BACKGROUND

Tests of "number facility" have been employed in intelligence testing (e.g., Wechsler, 1958), psychopharmacology (e.g., Crowell and Ketchum, 1967; Ketchum et al., 1973; and Michelson, 1961), behavioral toxicology (e.g., Johnson and Anger, 1983), and as a technique for testing and developing theories of human memory (e.g., Hitch, 1978).

This UTC-PAB test involves multidigit addition problems, the solution of which involves several cognitive structures as well as the utilization of cognitive procedures. For example, the subject must retrieve math facts from long term memory, retain intermediate results, keep track of carry and place information, and execute procedural knowledge (e.g., add units first, tens second, etc.). Therefore, the solution of these problems involves the retrieval of information from long term memory and working memory capacity in the form of short term storage and the execution of cognitive procedures. Figure 4 shows a model for the series of steps involved in the solution of these two column addition problems. This characterization assumes that subjects perform the operations from right to left; however, different strategies (e.g., solving the problem from left to right) and combinations of strategies have been used by subjects in the solution of multidigit addition problems (e.g., Hitch, 1978).

Research by Hitch (1978) with multidigit addition problems (adding two 3-digit numbers, or adding a 2-digit number to a 3-digit number) found that errors in addition could be accounted for by the loss of interim information (intermediate results and carries) and initial information. In his studies, Hitch presented the math problems auditorially and subjects were not allowed to take notes. Therefore, the loss of initial information (the numbers presented for addition) accounted for a significant proportion of the errors in addition.

The UTC-PAB version of the test involves visual presentation of the math problems that remain visible until the subject begins to enter his/her answer. This will make the loss of initial information a negligible factor. This is especially true since the subject is to enter the most significant digit first which requires the solution of the entire problem. Thus, errors in calculation, for the UTC-PAB version, can be attributable to the loss of intermediate solutions and carry information.

The number of carries required in the solution of a multidigit addition problem has been shown to have an effect on solution times. For example, Hitch (1978) found that solution latencies were fastest for problems that did not require carrying (e.g., 434 + 51) and slowest for those that

```
Problem:       29
               32
               13
```

Calculation/Memory
Retrieval and Procedures

		Intermediate Result	Carry	Answer
1)	9+2 =	11	--	--
2)	11+3 =	14	--	--
3)	Set carry and store partial answer	--	1	4
4)	2+1 =	3	--	4
5)	3+3 =	6	--	4
6)	6+1 =	7	--	4
7)	Concatenate partial results			74
8)	Respond			

Figure 4.   Sequence of Steps Involved in the Solution of a
            Two-Column Addition Problem [This Characterization
            Assumes Addition of Each Separate Column in a Right
            to Left Order]   (Adapted from Hitch, 1978)

required carrying in both the tens and hundreds (e.g., 434 + 87). These
results are consistent with the suggestion that carrying is a separate
stage (i.e., separate from storing intermediate results) that requires
extra processing time.

An additional factor that will contribute to the solution latencies is the
speed with which subjects retrieve arithmetic information from long term
memory. Ashcraft and Stazyk (1981) examined subject's ability to verify
the truth value of simple addition problems (e.g., 7+1 = 8 versus
7+1 = 9). Single digit addition problems were presented with either a
correct or incorrect solution and subjects were required to answer "true"
or "false" by pressing one of two buttons. True problems were generally
responded to more quickly than false problems. Furthermore, for false
problems it was found that the greater the difference between the stated
and the correct solution, the faster the response. Finally, an experiment
involving complex addition (14+12 = 26) indicated that subjects solve these
problems in a series of elementary steps.

Ashcraft and Stazyk (1981) interpreted their results in terms of network
models of semantic memory (e.g., Collins and Loftus, 1975). That is, for
adults simple mental addition is largely a memory retrieval phenomenon.
They appear to rely on a stored systematic structure of knowledge and not
on such procedures as counting.

Research by Winkelman and Schmidt (1974) also supports the memory retrieval
interpretation of simple mental arithmetic. Winkelman and Schmidt pre-
sented subjects addition and multiplication problems with either a correct
or incorrect solution. Each problem was presented separately and the sub-
ject's task was to respond true or false as quickly and as accurately as
possible. The reaction times for associative confusion problems (e.g.,
7+2 = 14 or 7x2 = 9) were significantly slower than for the nonassociative
confusion problems (e.g., 7+2 = 8 or 7x2 = 13). This was interpreted to
mean that the problems were solved via a memory retrieval and that addition
and multiplication information is closely associated in memory. Similar
results have been found for addition and subtraction by Perez and Tracy
(1983).

In summary, this test appears to tap both long term memory and working memory capacity. Errors in computation will most likely result from the loss of carry or intermediate result information from working memory. The latency data will reflect the speed with which information is retrieved from long term memory and working memory processing and storage.

RELIABILITY

Reliability information on the UTC-PAB version of the test has not been located. However, Seales, Kennedy, and Bittner (1980) evaluated the reliability of a paper and pencil arithmetic test involving addition or subtraction of two 3-digit numbers, multiplication of two 2-digit numbers, and division of a 4-digit number by a 2-digit number. There were 18 subjects in the study who were tested on 15 consecutive days. A test consisted of 64 math problems during the first seven days and 96 problems for the remaining days. Subjects were tested in 10 minute sessions. Arithmetic performance (total attempted, total correct, and correct-minus-wrong) showed improvement over the first nine days of testing and remained stable thereafter. In addition, the interday correlations for the above three measures were relatively high (mean r = .935, .941, and .921, respectively).

The above results indicate that tests of simple arithmetic will yield relatively stable performance over time. However, it should be noted that the UTC-PAB version of the test differs from the above version in that it will involve only addition problems which will be presented one at a time on a CRT. If anything, the UTC-PAB version may prove to be more stable than the version tested by Seales et al. (1980) since such factors as operator confusion will be eliminated (see Winkelman and Schmidt, 1974). However, research that examines the reliability of this test needs to be conducted.

VALIDITY

This test appears to measure the construct of numerical ability (French, Ekstrom, and Price, 1963). As may be recalled, it was argued that the

UTC-PAB version of the test taps long term memory (e.g., math facts and strategies) as well as working memory capacity (storage of intermediate results and carries). Research by Ashcraft and Stazyk (1981) with single digit addition problems has supported the hypothesis that adults solve simple addition problems (e.g., math facts) via a process of memory retrieval. Research by Hitch (1978) with multidigit addition problems showed that people perform relatively complex mental calculations in a series of elementary stages. Also, the number of carries required by the problem had a systematic effect on response latencies. Finally, Hitch's research indicated that errors in calculation could be attributed to the loss of initial and interim information held in working memory.

The above indicates that the UTC-PAB two column addition test measures a subject's general number facility. Furthermore, the problems are presented in such a manner that working memory capacity is also being tapped.

SENSITIVITY

Tests of mental addition have shown sensitivity to a range of toxic, drug, and environmental stressors. Table 9 shows a list of studies that examined the effects of toxic agents and drugs on mental calculations.

TABLE 9. LIST OF STUDIES

Reference	Drug or Toxic Substance	Reported Effect
Johnson et al., 1974*	Carbon monoxide	No
Knave et al., 1978*	Jet fuel mixture	Yes
Repko et al., 1975*	Inorganic lead	No
Repko et al., 1976*	Methyl chloride	Yes
Crowell and Ketchum, 1967	Scopolamine	Yes
Ketchum et al., 1973	Atropine	Yes
	Scopolamine	Yes
	Ditran	Yes
Michelson, 1961	Parpanite	Yes

* Cited in: Johnson, B.L. and Anger, W.K. Behavioral toxicology. In
            W. Rom (ed.), Environmental and Occupational Medicine.
            Boston: Little, Brown, and Co., 1983.

Toxic Agents

Performance on mental addition has differentiated the control (no exposure) from the experimental group with such agents as methyl chloride and jet fuel mixtures. However, significant differences between control and experimental conditions were not evident for such agents as carbon monoxide and inorganic lead. It should be noted that the Johnson et al. (1974) study involved 23 ppm CO exposure (COHb level of 4 percent) and performance decrements were only evident in a dual task condition. Furthermore, the study by Repko et al. (1975) examined occupational exposure to inorganic lead in auto battery industry. The levels of exposure in this work setting were very low (80 mg lead per liter of blood) and the effects of inorganic lead were only evident on a test of eye-hand coordination.

Drugs

Mental addition has also been shown to be sensitive to the effects of drugs. For example, Ketchum et al. (1973) found that mental arithmetic performance deteriorated when subjects were administered atropine, Ditran, or scopolamine. Furthermore, it was observed that hallucinations, disorientation, and incoherence consistently appeared whenever mathematical performance fell below 10 percent of baseline. The dose necessary to produce a decline in mathematical performance to below 10 percent in half the population was calculated by probit analysis to be 152 mcg/kg, 20 mcg/kg, and 100 mcg/kg for atropine, scopolamine, and Ditran respectively (Ketchum et al., 1973, p. 131). Decrement in mental arithmetic has also been found by Crowell and Ketchum (1967) with scopolamine, and by Michelson (1961) with parpanite.

Environmental Stressors

Mental addition has also been shown to be sensitive to the effects of sleep deprivation (Haslam, 1985; Rosa et al., 1985) and the physiological effects associated with underwater diving (e.g., Baddeley and Flemming, 1967).

TECHNICAL DESCRIPTION

The three number-pairs are generated pseudo-randomly from the digits 1 through 9. Zero is disallowed. The display consists of five lines. The first three lines are the number pairs, vertically aligned. The fourth line consists of four underline characters. The fifth contains a solid nonblinking cursor located under the left most underline character. The display colors are white characters on a light blue background with a dark blue border.

Valid response keys are the digits 0 through 9, back space, and return (enter). Digits are echoed to the screen as entered. Invalid keys (e.g., letters symbols) are not echoed, but are tallied as "extras." Back space moves the cursor to the left, up to but not beyond the left-most digit's location, to allow overstrike correction. Each occurrence of back space is tallied as a "correction." The cursor moves to the right with each digit entry unless the maximum of four digits is already being displayed, in which case it remains in place awaiting back space, overstrike, or return.

Trial Specifications

Each trial consists of the following steps: (a) a math problem is presented in the center of the CRT; (b) as soon as the subject enters a valid response or 15 seconds have elapsed the problem disappears; (c) the subject enters the rest of the answer and presses the enter or return key; (d) the screen blanks for 500 msec or feedback is presented if the practice trials are being run; and (e) a new problem is presented. The subject has 15 seconds to enter the entire answer to a problem and is presented with an auditory signal (e.g., a "beep") if the deadline has elapsed. Furthermore, during training trials the length of the interstimulus interval is subject paced.

DATA SPECIFICATIONS

Each trial generates the following three dependent measures: (a) RT(1): This is the reaction time for the subjects first valid (digit) response;

that is, the left most digit in the answer. (b) RT(2): This is the
reaction time for pressing the return or enter key. The return or enter
key is pressed after the subject has entered the entire answer to the
problem. (c) Response Code: The response code indicates whether the
response was correct, incorrect, or terminated by the deadline. If the
deadline value elapses before the return key is pressed then RT(2) is set
to the value of the deadline. If the deadline elapses before any valid key
is pressed then RT(1) and RT(2) are both set to the value of the deadline.

The following summary statistics will be determined: (1) test duration in
seconds, (2) number of trials completed, (3) number and percent correct,
(4) number of backspace corrections, (5) number of extras, (6) number of
deadline occurrences, and (7) averages and standard deviations for RT(1)
and RT(2) computed separately for correct and incorrect responses.

TRAINING REQUIREMENTS

Subjects should be initially introduced to this test by presenting them
with the instructions. Following the instructions the subjects should be
presented with a minimum of 10 practice trials. The practice trials will
differ from the experimental trials as follows: (1) following each
response, the problem will be redisplayed with the correct solution along
with the response entered by the subject and the values for RT(1) and
RT(2), (2) this feedback will remain on the screen until the subject
presses a key; that is, for the practice trials the interstimulus interval
will be subject paced.

During the practice trials the experimenter should carefully evaluate the
subject's performance in order to determine that the instructions are being
followed. For example, the instructions stress that subjects respond
quickly and accurately; however, subjects may be sacrificing accuracy for
the sake of speed or, alternatively, they may be reaching the response
deadline too frequently. Furthermore, the experimenter should ensure that
subjects are entering the answers in the prescribed manner (e.g., from left
to right). It should be noted that one normally answers addition problems
from right to left.

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

This test examines your ability to perform mathematical calculations. The computer will present you with two column addition problems that you are to add as rapidly as possible. The answer must be given by entering the left hand digit first (usually the hundreds' or tens' digit) followed by the remaining digits. Once you make an entry the math problem will disappear. Therefore, it is very important that you know the entire answer to the problem before making an entry. If you make a mistake you can use the back space key to correct it. When you are satisfied with your answer, press the return key.

Example: 29
            32
            13
           ——
            7

Here you would press the 7-key; then the 4-key; then the return key. Remember to work as quickly and as accurately as possible. If you fail to respond in 15 seconds the problem will disappear and a new problem will be shown.

Section 6
MATHEMATICAL PROCESSING TASK (UTC-PAB TEST NO. 5)
(NUMBER FACILITY/GENERAL REASONING)

PURPOSE

The purpose of this self-paced mental arithmetic task is to test a subject's information processing resources associated with working memory. Specifically, the subject is required to: (a) retrieve information from long term memory, (b) update information in working memory, (c) sequentially execute different arithmetic operations, and (d) perform numeric comparisons.

DESCRIPTION

This test requires subjects to perform one or more addition and/or subtraction operations on single digit numbers and determine whether the answer is greater or less than five. Problems are presented in the center of a CRT screen in a horizontal format for left to right solution and are followed by an equal sign. The two possible responses for this task (greater than or less than) are entered via a two button keypad.

Three versions of this task are available for selection and are designed to produce significantly different response time performance. Each version requires three minutes of continuous performance by the subject and reaction times are recorded from onset of the problem presentation to the onset of the subject's response via the keypad. The three versions of this test are as follows: (a) low demand version--problems containing one mathematical operation, (b) moderate demand version--problems containing two mathematical operations, and (c) high demand version--problems containing three mathematical operations.

BACKGROUND

The present test was developed by Shingledecker (1984) and requires the execution of one, two, or three mathematical operations (addition and/or

subtraction) within a given problem. There is extensive literature regarding the solution of single operation problems; however, very little research has been directed at understanding the processes that underlie the solution of multioperation problems.

The present discussion will review the literature with regard to multi-operation problems. In addition, research dealing with single digit addition, single digit subtraction, and multidigit addition will be presented. The review of the mental arithmetic literature will involve a discussion of the four cognitive procedures identified in the PURPOSE section of this report (e.g., retrieval of arithmetic information, updating information in working memory, sequential execution of arithmetic operations, and numeric comparisons).

Chiles, Alluisi, and Adams (1968) developed a mathematical processing task that required the execution of two different mathematical operations (addition or subtraction). This task was designed to be used in the assessment of "mental workload." In addition, this task was incorporated into the multiple task performance battery (MTPB) which includes other information processing tasks (e.g., auditory vigilance, warning lights, meter monitoring, problem solving, choice reaction time task, tracking, and pattern discrimination). Research with this mental arithmetic task has examined subject's ability to time share among several tasks (e.g., Chiles and Alluisi, 1979; Chiles, Brani and Lewis, 1969; Chiles and Jennings, 1970; Hall, Passey and Meighan, 1965).

Research by Perez (1982) examined working memory storage and processing in the solution of multioperation problems. There were five experiments that examined response latency and error data for problems involving three operations (combinations of addition and subtraction). The arithmetic notation (e.g., algebraic notation, reverse polish notation) for the problems was varied in order to examine a subject's ability to manipulate arithmetic information. The results showed that: (a) errors in computation were a function of loss of operand information (the digits) and confusion between operations (e.g., adding instead of subtracting); (b) response latency was a function of the number of different operations in a problem (e.g., +-+

was slower than +++); and (c) an arithmetic notation such as reverse
polish, which minimizes transient memory load, led to better performance
relative to algebraic notation (the superiority of reverse polish notation
over algebraic notation was seen after very little practice with this
"unusual" notation).

Wanner and Shiner (1976) have also employed multioperation problems in the
study of working memory. Their experiment focused on the transient memory
load imposed by problems involving two operations of subtraction and paren-
thesis arranged in one of two different sequences: (a) left parenthesis
problems--(5-4)-1 or (b) right parenthesis problems--5-(4-1). The problems
were presented in left to right order on a CRT and were interrupted at
various points by the presentation of a series of words that were to be
recalled at a later time. The subjects solved the problems or recalled the
words at the end of the problem presentation (word recall and problem
solution occurred equally often over a series of trials).

Wanner and Shiner found that errors on the memory task and the math task
were related to the transient memory load imposed by pending operations.
For example, the transient memory load for the right parenthesis problems
is greater than for the left parenthesis problems since subjects will need
to wait until the entire problem is presented before computations can
begin.

Finally, research by Shingledecker (1984) employed multioperation problems
in order to generate a "mathematical reasoning" task with three levels that
produced reliably different performance. Figure 5 shows average reaction
time and subjective ratings of difficulty for the three levels of task
demand (these data are based on a sample of six subjects).

Shingledecker (1984) developed the present version of the task with the
following two considerations: (a) The task was developed as a standardized
loading task designed to place variable demands upon information processing
resources associated with the comparison of numeric stimuli. The selection
of the three task demand levels was determined empirically. That is, the
number of operations and combinations of the operations of addition and

63
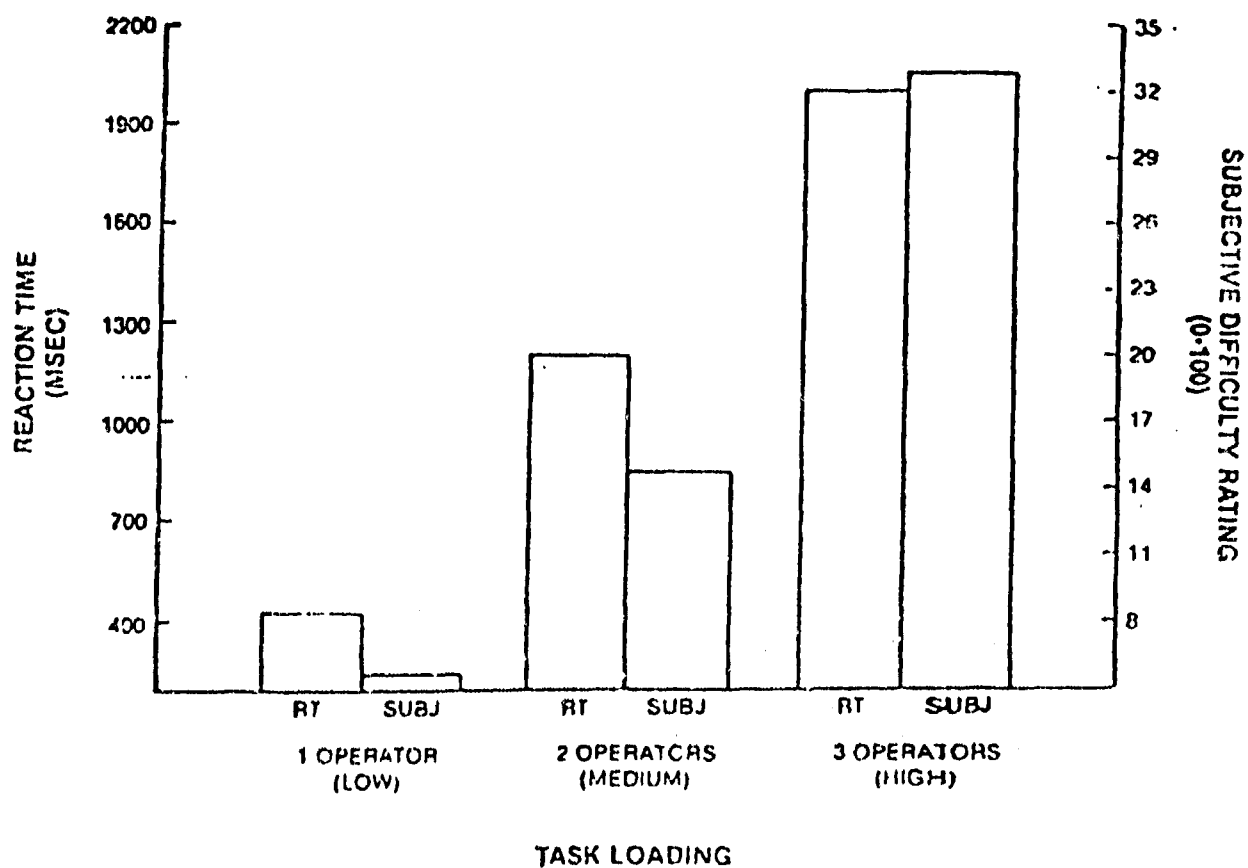
Figure 5. Mean Reaction Times and Subjective Difficulty Ratings for Mathematical Processing Conditions

subtraction were factorially combined during the initial phase of task development. The present levels were those that were shown to statistically differ from each other and represented an increasing degree of task difficulty (e.g., systematic increases in reaction time and number of errors). (b) The math processing task was based on a theoretical model of human information processing which posits three primary stages of processing and associated resources dedicated to perceptual input, central processing, and motor output or response activities (e.g., Wickens, 1984). The present task is presumed to tap resources that are primarily associated with central processing. Furthermore, this task involves relatively basic central processing activities such as information manipulation or transpositions based on implicit or memorized rules.

As described above, performance on the task may be broken down into four processing stages: (a) retrieval of arithmetic information from long term memory, (b) updating information in working memory, (c) sequential execution of different arithmetic operations, and (d) a numeric comparison. Literature regarding the above cognitive functions will be briefly outlined and discussed with respect to the three different versions of this test.

All conditions in the present task will require the retrieval of arithmetic information from long term memory. Research by Ashcraft and Battaglia (1978) (Ashcraft and Stazyk, 1981; Stazyk, Ashcraft and Haman, 1982) has shown that adults solve simple arithmetic problems via a memory retrieval. Adults appear to rely on a well organized memory structure and not so much on procedures such as counting. The data indicates that adults may have stored something analogous to "math tables" in long term memory.

The conditions involving multiple operations will require subjects to rapidly and sequentially carry out different arithmetic operations. Also, subjects will need to maintain and update an answer to the problem (e.g., "7+2 -4 -3", will result in the sequence of answers: 9, 5, 2). This type of activity will require working memory storage (e.g., Wanner and Shiner, 1976) and processing. Previous research (e.g., Perez, 1982) has shown that transitions from one operation to another (+ then -) requires more time

than sequential operations with the same operator (+ then +). The above suggests a memory priming effect in terms of arithmetic operations.

Finally, the present test will require subjects to compare an internally generated answer against a standard (is the computed answer greater or less than 5). Restle (1970) required subjects to compare the sum of two numbers (A + B) against a standard (C) and select the greater of the two (A + B or C). Results indicated that response latency decreased as the relative difference between the sum and the standard increased. The results were interpreted in terms of an analog operation in which subjects placed the magnitudes (A + B and C) symbolized by numbers on the number line for mapping and judging.

In summary, the UTC-PAB math processing test contains three versions that impose different demands on the human information processing system with respect to memory retrieval, updating working memory storage, sequential execution of mathematical operations, and numeric comparison. The three versions can be summarized in terms of the above four processing components: (a) Low Demand Version: The response latency will be a function of memory retrieval and number comparison. Errors may be due to associative confusion between operations (Winkleman and Schmidt, 1974), or the number comparison process. (b) Moderate Demand Version: The response latency will be a function of memory retrieval, updating working memory, serial execution of operations and number comparison. It should be noted that these problems may require two or more memory retrievals. For example, a problem such as "9+8 -5" will generate a value of "17" as the first result. The second calculation (17-5) may be performed in two stages (e.g., 7-5 = 2, 2+10 = 12) and, thus, the entire problem may require three memory retrievals of math facts (see Hitch, 1978, for data suggesting that adults solve "complex" math problems in a series of elementary stages). Errors in performance may result from a failure in one (or more) of the above four processing components. (c) High Demand Version: The response latency will be a function of memory retrieval, updating working memory, serial execution of operations and number comparison as with two operator problems; however, additional processing will be required with respect to

66

memory retrieval, updating working memory, and the serial execution of operations.

As can be seen, this test contains three versions that differ in terms of the degree to which subjects manipulate arithmetic information. Performance in this task appears to be diagnostic of long term memory retrieval and working memory storage and processing. For example, if a manipulation (e.g., drug) impairs performance on the two or three operation problems but not on one operation problem, one may conclude that the factor under study (e.g., drug) affects the manipulation of information in working memory but not the retrieval of information from long term memory.

RELIABILITY

Reliability information is not available for the UTC-PAB version of the mathematical processing test. However, reliability data have been obtained on a paper and pencil arithmetic test involving addition or subtraction of two 3-digit numbers, multiplication of two 2-digit numbers, and division of a 4-digit number by a 2-digit number (Seales, Kennedy, and Bittner, 1980). There were 18 subjects in this study who were tested on 15 consecutive days. A test consisted of 64 math problems during the first seven days and 96 problems for the remaining days. Subjects were tested in 10 minute sessions. Arithmetic performance (total, attempted, total correct, and correct minus wrong) showed improvement over the first nine days of testing and remained stable thereafter. In addition, the interday correlations for the above three measures were relatively high (mean r = .935, .941, and .921, respectively).

Chiles, Jennings, and Alluisi (1978) reported reliability coefficients for a multioperation task which required the addition of two 2-digit numbers and the subtraction of a third 2-digit number (e.g., 12+15 -13 =). There were 94 subjects in this study; however, only 51 subjects were tested on two consecutive days. The subjects received 15 minutes of practice before the start of testing. The math task was performed in conjunction with one of the following two tasks; a problem solving task and a manual tracking task. Also, the subjects were required to perform two monitoring tasks

(light and meter monitoring) in addition to the above tasks. The authors computed reliability coefficients by correlating performance on the math task across all of the task combinations. The average correlations for those subjects tested for one day were .73 and .82 for solution time and accuracy, respectively. For those subjects tested on two consecutive days, the average correlations were .01 and .71 for solution time and accuracy, respectively. The above reliability data indicate that performance on a math task is relatively stable over time. Furthermore, research by Chiles et al. (1978) indicates that performance on multioperation problems is reliable for both speed of solution and accuracy. However, the present test contains three different versions that differ from the studies reviewed here. The present data suggests that the UTC-PAB mathematical processing test will yield stable performance over time (even with little practice); however, additional reliability data is needed.

VALIDITY

This test appears to tap resources associated with working memory storage and processing. In addition, the present test requires the retrieval of arithmetic information (e.g., math facts) from long term memory and a number comparison judgement. As stated earlier, research with single digit addition problems (e.g., Ashcraft and Stazyk, 1981) has supported the hypothesis that simple addition problems are solved via a memory retrieval process (this is true of adults). In addition, research with multidigit addition problems (e.g., Hitch, 1978) has shown that the solution of complex math problems require working memory storage and are solved in a series of elementary steps.

Research by Chiles et al. (1978) with multioperation problems also indicates that a math processing task taps resources associated with working memory processing. For example, in their study a multioperation arithmetic task was performed concurrently with either a problem solving task (e.g., code lock solving task) or a manual tracking task. Performance on the math task was worse when performed with the problem solving task (percent correct = 70.94) relative to when it was time shared with the tracking task (percent correct = 82.37). The manual tracking task appears to tap

resources associated with motor response processing and should not inter-
fere with a task such as the math processing test (e.g., Wickens, 1984).
On the other hand, two tasks that involve working memory processing (e.g.,
math task and the code lock solving task) do interfere with each other.

SENSITIVITY

This review indicates that the UTC-PAB mathematical processing task tests
resources associated with working memory storage and processing. Perform-
ance on this test is sensitive to the load imposed by a secondary task
which involves working memory processing (e.g., code lock solving task).
This selective sensitivity to secondary task load suggests that the mathe-
matical processing task has a utility as a diagnostic tool.

The present version of this test has not been employed in the study on the
effects of toxic substances, drugs, or environmental stress. However, the
multioperation task developed by Chiles et al. (1968) has been employed in
behavioral toxicology research. For example, Morgan and Repko (1974)
tested 316 workers manufacturing auto storage batteries for 3 to 16 years
and a control group of 112 workers. The purpose of this study was to pro-
vide a quantitative assessment of change in performance which could result
from occupational exposure to inorganic lead. The study did not reveal a
significant difference in mathematical processing performance between the
lead exposed workers and the control group. Furthermore, the only dif-
ference in performance between the lead exposed and control subjects was on
a test of eye-hand coordination (the exposed workers had less than 80 mg of
lead per liter of blood which is a relatively low level of lead expo-
sure). The exposed workers were slower than the control workers on the
test of eye-hand coordination.

Negative results have also been demonstrated by Chiles and Jennings (1970)
with respect to the effect of alcohol consumption on mathematical pro-
cessing. This study involved several other tasks from the MTPB (e.g.,
warning lights, problem solving, etc.) which were performed concurrently
with the math processing task. However, research by Repko et al. (1976)
which studied the effects of exposure to methyl chloride on human

information processing found a difference between exposed and control subjects on mathematical processing. This study involved 45 control subjects and 122 workers exposed to approximately 35 ppm of methyl chloride.

The above indicates that the mathematical processing task developed by Chiles et al. (1968) is sensitive to secondary task load if the secondary task requires working memory storage and processing. In addition, toxicology research by Repko et al. (1976) showed that the math processing task was sensitive to the effects of methyl chloride exposure. However. research involving exposure to lead (Morgan and Repko, 1970) and alcohol consumption (e.g., Chiles and Jennings, 1970) did not show significant performance decrements on the math processing task.

The present data suggests that the UTC-PAB math processing task may be sensitive to the effects of drug if the drug disrupts working memory processing. However, relatively little research has been conducted with the UTC-PAB version of this test. Furthermore, the UTC-PAB version of this test contains three versions that appear to differ qualitatively with respect to cognitive operations. That is, the low demand version appears to principally involve retrieval of math facts from long term memory and a math comparison process; however, the moderate and high demand versions appear to involve working memory storage and processing in addition to long term memory retrieval and the number comparison process.

TECHNICAL DESCRIPTION

Problems are presented in the center of the CRT in a horizontal format for left to right solution and are followed by an "equal" sign. Problems are randomly generated with the following restrictions: (1) the digits 1 through 9 are used, (2) the correct answer may be any digit from 1 to 9 except 5, (3) half of the problems presented in a set of trials will have an answer greater than 5, the other half will have an answer less than 5, (4) when problems are solved from left to right, cumulative intermediate totals must have a positive value, and (5) no problems will contain the same digit twice unless they are both preceded by the same operator (e.g.,

70

+6 and -6 would not appear in the same problem). Example problems are shown on Figure 6.

The subject responds to each problem by pressing one of two keys on a keypad in order to indicate whether the answer to the problem was greater (>) or less (<) than 5. The nature of the manual response requirements is the same for the low, moderate, and high demand versions of the test.

Trial Specifications

Each trial will consist of the following steps: (a) a math problem will be presented in the center of the CRT; (b) as soon as the subject enters a valid response or the deadline has elapsed (1.5 seconds for low demand, 3 seconds for moderate demand, and 4 seconds for the high demand version) the problem will disappear; (c) the screen blanks for 500 msec or feedback is presented if the practice trials are being run; and (d) a new problem is presented. During the experimental trials subjects are tested continuously for 3 minutes with above procedure.

DATA SPECIFICATIONS

For each 3-minute trial block the following data will be recorded: (a) Reaction Time (RT): Reaction times will be recorded for each response (e.g., > or <) in the trial block. (b) Response Code: The response code will indicate whether the response was correct, incorrect, or terminated by a deadline. If the deadline value elapses before the key press then the RT for that trial will be set to the value of the deadline.

The following summary statistics will be determined: (1) number and percent correct, (2) number of deadline occurrences, and (3) averages and standard deviations for RT computed separately for correct and incorrect responses (incorrect responses resulting from deadline termination will not be considered).

A. Low Demand Version   Correct Answer

     7 - 1 =   > 5
     4 + 2 =   < 5
     9 - 7 =   < 5

B. Moderate Demand Version

     6 - 5 + 2 =   < 5
     9 - 1 - 2 =   > 5
     2 + 1 - 1 =   < 5

C. High Demand Version

     9 + 7 - 5 - 1 =   > 5
     1 + 1 + 3 - 1 =   < 5
     8 - 7 + 5 - 3 =   ↓   < 5

Figure 6. Examples of Low, Moderate, and High Demand Problems [For
           Each Problem Subjects will Depress One of Two Leys in Order
           to Indicate Whether the Answer was Greater (>) or Less (<)
           than 5]

72

TRAINING REQUIREMENTS

Subjects should be initially introduced to this test by presenting them with the instructions. Following the instructions the subjects should be presented with a minimum of 10 practice trials. The practice trials will differ from the experimental trials as follows: (1) following each response, the problem will be redisplayed with the correct solution along with the response entered by the subject and the value of the RT, (2) this feedback will remain on the screen until the subject presses a key; that is, for the practice trials the interstimulus interval will be subject paced.

During the practice trials the experimenter should carefully evaluate the subject's performance in order to determine that the instructions are being followed. For example, the instructions stress that the subject respond "quickly and accurately"; however, subjects may be sacrificing accuracy for the sake of speed or alternatively they may be reaching the response dead-line too frequently. Furthermore, the experimenter should stress the fact that problems should be solved in a left to right format in order to avoid negative intermediate results.

For this task, training times required for subjects to reach asymptotic performance have been determined. For example, training times for the three test versions are as follows: (1) low demand version--seven 3-minute trials; (2) moderate demand version--10 to 14 3-minute trials; and (3) high demand version--10 to 30 3-minute trials. It should be noted that the above training times are based on one study that utilized a rather small sample size (N = 6). In addition, the above subjects were from a subject pool and were highly practiced on behavioral performance tasks.

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

In the Math Processing task, you must solve a number of simple addition and subtraction problems to determine whether the correct answer is greater or less than 5. The two possible reponses on the task are "greater than" (>) and "less than" (<). Greater than responses are entered on the rightmost key and less than responses on the leftmost key. No problem will ever have the value 5 as the correct answer.

You start the task whenever you are ready by pressing any of the response keys. Testing periods last for 3 minutes each. Math problems appear one at a time on the screen, and should be solved from left to right. Always perform the additions and subtractions in the order that they appear in the problems. As soon as you respond to a problem, a new problem will appear. Try to perform the task as quickly and accurately as possible. Go as fast as you can, but if you start to make errors because you are trying to go too fast, slow down. You should try to respond correctly to every problem. At the end of the 3-minute testing period, the task will automatically stop and the screen will go blank.

The number of additions and subtractions to be performed in each problem will vary from one 3-minute period to another. On some periods problems will require only one addition or subtraction to be performed; on others, two additions and/or subtractions; and on others, three operations. However, in a given 3-minute test period, all problems will have the same number of mathematical operations.

Section 7
CONTINUOUS RECOGNITION TASK (UTC-PAB TEST NO. 6)
(WORKING MEMORY--ENCODING and RECOGNITION)

PURPOSE

The Continuous Recognition Task is designed to place variable demands upon processing resources associated with encoding and storage in working memory. The task tests a subject's ability to encode, rehearse, recall, and compare numbers in short term memory on a continuous basis.

DESCRIPTION

The memory test consists of a random series of visual presentations of numbers which the subject must encode in a sequential fashion. As each number in the series is presented for encoding, a probe number is presented simultaneously. The operator must compare this probe number to a previously presented item at a prespecified number of positions back in the series. The operator must decide if the previously presented item is the same as or different from the probe number. Thus, the task exercises working memory functions by requiring operators to accurately maintain, update, and access a store of information on a continuous basis. Task difficulty is manipulated by varying the number of digits which comprise each item, and the length of the series which must be maintained in memory in order to respond to recall probes.

BACKGROUND

The Continuous Recall Task is a test of running working memory. Running memory involves the short term memory of symbols under a continuously changing storage state. That is, items are presented in an unsystematic running order and require the continuous recall of a recent item for each successively presented item. Once an item has been recalled it is excluded from short term memory while the current item is encoded. The task involves mental processes similar to those used in the monitoring of

instrument gauges, where the retention and recall of only recent occurrences are appropriate for efficiency while the exclusion of past items is necessary.

The early predecessor of the Continuous Recall Task was the Running Matching Memory (RMM) task. The RMM task was devised by Moore and Ross (1963) in order to investigate the effects of context on running memory. The RMM task requires the subject to say whether each successively viewed symbol in a running, randomly ordered series was the "same" or "different" from the symbol seen in a specified number of symbols back in the series. For example, a 2-back match would involve comparing the third symbol presented with the first, fourth with the second, fifth with the third, and so on for 40 trials. Moore and Ross (1963) used 2-back sequences and manipulated context by varying: (1) the number of different symbols comprising individual series of symbols (+, -, , 0), and (2) the different symbol combinations occurring within a symbol series. The task was subject paced and the mean number of errors was measured.

Different combinations of preceding 3-symbol sequences (e.g., ++-, -+-, +++) were analyzed for each number of symbols (two, three, or four) comprising the series. Results showed that when more symbols were used, mean number of errors declined. Also, mean errors declined when the exposed symbol was "novel" (unrelated to symbols already in memory).

The RMM task was also used to investigate serial order as a unique source of error in running memory (Ross, 1966a). Task difficulty was varied by having subjects perform 2-back, 3-back, 4-back series, or some combination of two series. The time allowed for recall was also manipulated. Total symbol processing time was either 2.75 seconds or 5 seconds. Results revealed a constant amount of error for the "XYY" symbol combination (i.e., -++ or +--) regardless of symbol processing time and retained symbol load. These results indicate that memory for serial order produces unique sources of error in running memory. Total errors increased as the number back to be recalled increased, and as total processing time decreased.

Ross (1966b) also devised a two channel version of the RMM task. In this version of the task subjects were required to perform 1-back matches on symbols viewed on the left display, and at the same time perform 2-back matches on different symbols presented on the display on the right. The order in which the two displays were responded to was varied. It was found that those subjects who performed a 1-back match before performing a 2-back match committed more errors on 2-back matching than subjects who performed 2-back matches each trial. Serial order (symbol combinations) was a greater source of error than was symbol load.

The Continuous Recognition Task was also implemented by Hunter (1975) as part of an Air Force Psychomotor/Perceptual Battery. The task involved both immediate and short term memory of symbols under continuously changing storage state. This version of the test consisted of a continuous random series of presentations of one of nine geometric keyboard figures. The subject was instructed to depress the appropriate keyboard button for the figure which appeared two figures back when the third figure appears on the display. Each time a new figure appeared, the subject was to press the appropriate button for the figure which appeared 2-back. In the immediate memory test the figures were displayed for a two second stimulus duration with a two second intersignal interval. The delayed memory portion of the test had an intersignal interval of 5 seconds. For both parts, the number of correct responses was taken as the dependent measure. The performance data indicated that subjects performed better in the delayed memory condition than in the immediate memory condition. Subjects averaged 16.06 correct responses out of 25 stimuli in the delayed condition, but averaged only 12.77 correct responses in the immediate condition. A factor analysis was performed for all the tests in the battery. The Continuous Recall Task obtained a high loading on one of the factors identified as "figural memory." This factor principally defined those variables in which the subject must remember strings of geometric figures in particular order.

The UTC-PAB version of the Continuous Recognition Task is taken from the Criterion Task Set version (Shingledecker, 1984). The Criterion Task Set (CTS) is a battery of standardized cognitive tasks designed to place variable demands on resource allocation for a variety of cognitive

processes. In the CTS version of the Continuous Recognition Task, a random series of numbers is visually presented in a sequential fashion which the subject must encode. As each number in the series is presented, a probe number is presented simultaneously above it. The operator must compare this probe number to a previously presented item at a prespecified number of positions back in the series. Once the operator has made the appropriate recall, he/she must decide if that item is the same as/or different from the probe number. Three significantly different task demand levels are produced by the following conditions: low demand--recalling one position back one digit number; medium demand recalling two positions back two digit numbers; high demand--recalling three positions back four digit numbers. The task is subject paced and roughly half of the probe numbers result in a recall comparison of "same." Reaction time and subjective difficulty measures show significant differences between the three levels of difficulty. Reaction times averaged approximately 575, 750, and 1200 mseconds for low, medium, and high demand respectively.

The Continuous Recognition Task has not been formally related to any specific model of memory. Hunter (1975) states only that the task involves both immediate and short term memory of symbols under continuously changing storage state. It is evident from the nature of the task that different processing resources associated with short term (working) memory are required. The subject must encode items into working memory and maintain the items in memory by rehearsal. The order of the items in memory must also be maintained. As each subsequent stimulus is presented, the subject must recall one of the items in memory, compare it to the newly presented item, make the appropriate response, and encode the new item for rehearsal. This process is repeated on a continuous basis. The rationale for requiring subjects to make "same" and "different" comparisons was that it necessitated subjects to perceive and make use of every symbol before it was placed in their memory store. Thus, retention errors owing to subject's failing to perceive the symbols should be minimized. This is not the case in the UTC-PAB version of the task. In this version, probe items do not become target items. The new target item is displayed below the probe and, thus, is not processed before it is encoded into short term memory. Requiring a match to be made of each symbol exposure also cut down on

any tendency for the subject to categorize symbols according to serial patterns.

Symbol processing time, retained symbol load, and serial order of symbols all uniquely influence performance on the task. The data from Ross (1966a) show a large decrease in mean errors on the task, for each level of processing load, when symbol processing time was increased from 2.75 seconds to 5 seconds per symbol. Hunter (1975) also found a significant difference between delayed and immediate recall. The increase in performance on the task in the delayed condition further suggests that the rehearsal of items is important to the task. In most short term memory experiments, the likelihood that an item in a list will be recalled tends to increase with the amount of time available for its rehearsal. The extended interstimulus interval in the delayed condition allows for more rehearsal of the stimuli than in the immediate recall condition.

Retained symbol load is a function of the number of stimulus items in memory due to the number of items back the subject is to recall and/or the complexity of each stimulus item. Subjects performing longer match backs have to retain more symbols on every trial causing an increase in average storage load. If the items to be retained are large (e.g., 4-digit numbers versus 1-digit numbers), the average storage load would be further increased. The experiments conducted by Ross (1966a) and Shingledecker (1984) both show significant increases in errors with an increase in retained symbol load. Also, if symbol processing time was increased by an increment proportional to the increase in match back length, one second per additional symbol, longer watch backs would still tend to produce more errors (Ross, 1966a).

The experiments utilizing the RMM task (Moore and Ross, 1963; Ross, 1966a; Ross, 1966b) have demonstrated that serial order of symbols affects running memory. Certain symbol combinations that have been encoded into short term memory produce more errors than other combinations. This finding holds for different symbol processing times and different length match backs. The serial order effect seems to occur when there is immediate repetition of an item; that is, when the required item to recall is the same as another item

in memory. Thus, the greater number of possible items, the less chance there is of a serial order effect. Also, the effect does not occur with 1-back matching as there is no other item in memory besides the one being recalled. The serial order effects discovered indicate that the processes necessary to retain the order of the items may be uniquely different from other processes involved in running memory. Serial order effects should not have an influence in the UTC-PAB version of the task, however, since a large number of symbols are used.

RELIABILITY

Reliability estimates for the Continuous Recognition Task were computed by Hunter (1975) using the Kuder-Richardson Formula-20 (KR-20). Computations were based on performance data (percent correct) from 305 subjects. Reliability for both the immediate and delayed memory versions of the task was .93. This type of reliability (interitem consistency) is based on the consistency of responses to all items in the test. The interitem consistency is found from a single administration of a single test. No studies have reported test-retest reliability for the Continuous Recall Task. Test-retest reliability involves computing the correlation between scores obtained by the same person on two or more administrations of the test. Since the UTC-PAB version of the task is intended for use in environmental studies, which usually require repeated testing of subjects, test-retest reliability would be more beneficial than interitem consistency for this task. Thus, experiments utilizing repeated testing of the Continuous Recall Task that report test-retest reliability would be of great value to this task.

VALIDITY

The Continuous Recognition Task is intended to test processing resources associated with short term memory by requiring subjects to encode, rehearse, retrieve, and compare numbers in running memory on a continuous basis. Since the serial order of items is not predictable, and good performance requires continuous discarding of items that are no longer

80

useful, the Continuous Recognition Task is closer than list learning tasks
to everyday information processing.

A factor analysis revealed that the Continuous Recognition Task was highly
loaded (.85) on a factor involving the memory of strings of figures in a
particular order (Hunter, 1975). Construct validity is further supported
by the replication of several results in a number of experiments. Longer
symbol processing times have been shown to increase performance on the task
(Hunter, 1975; Ross 1966a). This result is consistent with current the-
ories of short term memory rehearsal (Craik and Watkins, 1973). Ross
(1966a) and Shingledecker (1984) have demonstrated that larger retained
symbol loads on this task result in an increase in the number of errors.
Also, Moore and Ross (1963), and Ross (1966a,b) demonstrated the serial
order of lists results in an increase in the number of errors.

SENSITIVITY

The Continuous Recognition Task has not been extensively used in environ-
mental research. The only such research reported utilized the two channel
RMM task to assess the effects of transverse G-stress on short term memory
(Ross and Chambers, 1967). In one earlier study the 2-channel RMM task was
found to be differentially sensitive to a range of alcohol dosages
(Carpenter and Ross, 1965). However, the action of G-stress provides a
sharp contrast with that of alcohol in that a constant physical force is
applied for an exactly specified period of time.

Ross and Chambers (1967) designed an experiment to determine the effect of
different amounts of G-stress on RMM performance. The investigators were
also interested in determining whether previously found serial order
effects would be manifest under G-stress. The RMM task involved the random
presentation of the numbers one and two in the left display and the random
presentation of the signs "+" and "-" in the right display. Symbols in the
two displays came on simultaneously for 2 seconds and went off simultane-
ously for .75 seconds allowing a total information processing time of
2.75 seconds. The viewed number on the left was matched with the previous
(1-back) number as to whether it was "same" or "different," while the

viewed sign on the right was matched as to whether it was "same" or "different" from the next to the last (2-back) sign. Subjects responded on each successive trial by twice pressing the response buttons on a four button response handle.

Subjects were administered G-stress under controlled conditions by use of a human centrifuge. Either 3, 5, 7, or 9 Gs were induced in a given 2-minute and 18-second experimental run. Only transverse stress (chest to back) was induced. Subjects performed the RMM task under each stress level and in a static state (1-G lying on back) after each condition.

No memory deficit was found at 3-G. Significant memory deficit was found at 5-G and 7-G with still greater deficit at 9-G. Most of the deficit occurred during the latter half of each 2-minute and 18-second stress period. Performance decrements during dynamic (stress) series did not carry over to subsequent static series; therefore, the decrements were produced by the immediate situation rather than as a product of fatigue. Results also indicated that for retained symbols, serial order errors are not sensitive to G-stress. However, stress versus nonstress differences were found in serial orders that included a previously correct symbol that subjects had to discard. This finding led the authors to hypothesize that subjects under G-stress curtailed the number of symbols they processed during a memory match. That is, under G-stress the discarded symbol preceding the matched symbol is retained to a lesser extent. This curtailment of symbols is advantageous insofar as it lessens interference from the symbol that should be discarded. Such an improvement is, however, only relative, as total errors for all G-stress conditions were greater than for static conditions.

TECHNICAL DESCRIPTION

The UTC-PAB version of the Continuous Recognition Task contains three standard loading levels. In the low demand condition, memory items are one digit in length and subjects are required to recall one item back in the series. In the moderate demand condition, items are two digits long and recall is two positions back. In the high demand condition, items are four

digits long and recall is three positions back. In all conditions of the task, items are displayed serially on a CRT screen with the following restrictions: (1) test numbers must be randomly generated, (2) only the numerals 1-9 are used, and (3) roughly half of the probe numbers must result in a recall comparison of "same." Test numbers and probe numbers are simultaneously presented as well as terminated. The test numbers always appear below a line centered on the CRT while the probe numbers appear directly above the line. Since the probe number does not become a test number, each new test number is not preprocessed before it is encoded into memory.

Test trials consist of 3 minutes of continuous performance. In all conditions, the task is subject paced within the limits of selected deadline reaction times. Maximum acceptable reaction time in the training mode is 15 seconds for all conditions. If the subject does not respond within 15 seconds after the onset of the test item, the next item is automatically presented.

In the testing mode, the reaction time deadlines are reduced: 1.1 seconds for the 1-digit 1-back condition; 1.7 seconds for 2-digits 2-back; and 2.3 seconds for 4-digits 3-back. The probe and target display is approximately 1.25 inches high. Each number is approximately 2.5 inches by .13 inches, and should be viewed from a distance of roughly 60 cm. Responses are entered on a two-button keypad. A new display of numbers is presented whenever a button is pressed or when the deadline time has elapsed.

Trial Specifications

Each trial of the Continuous Recall Task lasts for 3 minutes. A trial is initiated by pressing either of the response keys. At this point, the first test and probe numbers are presented. The subject is to encode the test number and not process the probe number which shall be "00." The subject shall encode sequentially test numbers and ignore probe items until the number of presentations has equalled the number of match backs that the subject is to perform. For example, if the subject was to perform the

83

2-digits, two positions back condition, the subject would encode the first two test digits while ignoring the probe items. On the third presentation, the subject would begin comparing the probe items to the test items which occurred two positions back. The subject would continue responding on every subsequent presentation until the 3-minute period has expired.

DATA SPECIFICATIONS

Unprocessed data are collected and stored on all trials. The data to be recorded are: (1) time of onset of the probe and test item, (2) time of subject's response, (3) identity of test and probe numbers, and (4) identity of response. The following summary statistics will be calculated for each trial: (1) number of problems responded to, (2) number and percent correct, (3) number and percent of errors of commission (incorrect responses), (4) number and percent of errors of omission (no response within deadline), (5) number and percent of total errors, (6) mean and median RT, and (7) standard deviation of reaction time.

TRAINING REQUIREMENTS

A typical strategy suggests that subjects first inspect the probe number above the line and decode whether or not it matches the appropriate item in memory. Next, the test number below the line is encoded. Finally, the decision response is made on the key pad. Major practice effects for the Continuous Recall Task are eliminated within five to seven 3-minute training trials at each of the three loading levels.

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3.   Repeat the practice trials if it appears that the subjects require additional practice with the test.

4.   Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

In the Continuous Recognition Task, you will be presented with a series of two numbers, one appearing above the other. Both numbers will consist of either one, two, or four digits. Your task is to memorize the bottom number, and decide whether the top number is the same as the bottom number that you memorized one, two, or three screens earlier. In one task condition, the numbers will be single digits (1-9), and the top number must be compared to the bottom number from the previous screen (1-digit 1-back). When the numbers are composed of two digits (10-99) the top number is compared to the bottom number appearing two screens back (2-digits 2-back), and when the numbers are four digits long (1000-9999), the top number is compared to the bottom number that appeared three screens back (4-digits 3-back). For example, in the 1-digit 1-back condition, if the stimuli were:

Screen 1	Screen 2	Screen 3	Screen 4
$\frac{0}{4}$	$\frac{4}{7}$	$\frac{7}{2}$	$\frac{3}{1}$

the correct responses would be: Screen 1 either same or different (neither response is incorrect because there is nothing one screen back from the first screen; press either key when you have memorized the bottom number); Screen 2--"same," because the top number "4" matches the bottom number on the previous screen; Screen 3--"same," since the "7" on top is the same number as the bottom "7" on Screen 2; Screen 4--"different," because the "3" does not match the "2" on Screen 3. The procedure is the same in the other conditions except that considered responses are not required for the

first two or three screens and the top numbers are compared two or three screens back.

In order to successfully perform this task, you will have to do two things every time the screen changes. First you must memorize the bottom number, and then you must indicate whether the top number on the current screen is the same or different from the bottom number on one of the previous screens. Remember that you must memorize the bottom number before you respond, because a new screen will appear when you press a key, and the information will be lost. Also, keep in mind that in the 1-digit 1-back condition, the response to the first screen doesn't matter; likewise the first two or three responses do not matter in the 2-back 2-digit and 3-back 4-digit conditions respectively. On the first "memorization only" screens the top number will always be a zero.

You will be starting each data collection period by pressing either response key. Data collection trials last 3 minutes. You should try to respond as quickly and as accurately as possible. When you enter a response, the next screen will immediately be displayed. If you find your-self making errors from trying to go too fast, slow down. However, do not take any more time than is necessary to remember the bottom number and cor-rectly respond to the top number. At the end of the 3-minute period the task will stop and the screen will go blank.

Section 8
FOUR-CHOICE SERIAL REACTION TIME (UTC-PAB TEST NO. 7)
(ENCODING, CATEGORIZATION, RESPONSE SELECTION)

PURPOSE

This task is designed to evaluate information processing resources dedicated to stimulus encoding and categorization, and response selection, although it is probable that resources dedicated to encoding are tapped most heavily.

DESCRIPTION

A blinking "+" (plus sign) imposed on the cursor in one of four quadrants of a CRT is presented to the subject. The subject is instructed to press the key (one of four) on the keyboard that corresponds to the quadrant with the blinking "+." The blinking "+" remains in a quadrant until one of the four keys is pressed and then randomly reappears in any one of the quadrants. If none of the four buttons are pressed within 2.5 seconds, a bell rings at 0.1 second intervals until a response is made. Subjects are instructed to respond as quickly and accurately as possible. The task lasts 6 minutes.

BACKGROUND

Development of UTC-PAB Version of the Four-Choice Reaction Time Task

This task is a modification of the four-choice reaction time task developed by Wilkinson and Houghton (1975). The authors' objective was the field application of a classical laboratory paradigm. The achievement of this objective was realized as a result of the utilization of a battery operated tape recorder which created the potential for satisfaction of the two chief demands of field testing: self administration and portability. The tape recorder was adapted to perform the triple function of housing the display and response apparatus, generating a program of stimuli, and recording the

response data. The program generation and data storage capabilities were
made convenient utilizing standard magnetic tape cassettes.

The adaptation of the Wilkinson and Houghton (1975) portable four-choice
reaction time test to microcomputer administration (as per the UTC-PAB ver-
sion) was presented by Ryman, Naitoh, and Englund (1984). This adaptation
is especially useful with reference to the widespread availability and
efficiency of digital computers. A computer can perform all of the tasks
assigned by Wilkinson and Houghton (1975) to the portable tape recorder
more quickly and efficiently. Self administration of the task remains a
possibility with the computer version. The microcomputer adapatation may
not be as readily portable as a tape recorder, but computer technology is
certainly moving in this direction.

The Choice Reaction Time Paradigm: An Overview

Any choice reaction time experiment is usually characterized by the fol-
lowing properties: a set of possible stimuli, a set of possible responses,
and a mapping of the stimuli into the response that is specified by the
experimenter. On a given trial, one of the possible stimuli is presented
to the subject whose task consists of making the response appropriate for
this stimulus as quickly as possible (Smith, 1968). Of course, reaction
time is the major dependent variable, but this paradigm lends itself to
several others (Table 10).

The origin of this notion of applying reaction time measures to decision
making behavior must be attributed to the 19th century scientist,
F. C. Donders (1969; translated from the 1868 original) in his development
of the subtraction method. Utilizing this method, Donders attempted to
understand various "mental processes" by attempting to indirectly measure
the time required by a particular process. To summarize the logic underly-
ing the subtraction method: A reaction time task can involve any number of
mental processes. If such processes operate serially (which may actually
be a faulty assumption), then the reaction time required by a particular

TABLE 10. EXAMPLES OF RESPONSE MEASURES FOR THE CHOICE
REACTION TIME PARADIGM

MEASURE	REFERENCES
Reaction Time (RT) [Elapsed time between stimulus onset and response] Number of Responses Per Unit Time Number of Errors Number of Correct Responses Per Unit Time	No specific references are included for these traditional measures
Decrement of RT Within a Block of Trials [Mean RT for the first half of the block divided by the mean RT for the second half of the block]	Herbert et al., 1983
Coefficient of Variability	MacFlynn et al., 1984
Movement Time (MT) [Total response time (TT) minus decision time]	Krause and Bittner, 1982
Number of Gaps [Total number of response intervals of 1 second or more]	Wilkinson and Houghton, 1975
Number of Pauses [The Interresponse interval which exceeds the mean RT by a factor of 1.5]	MacFlynn et al., 1984

process can be assessed by comparing reaction times associated with different reaction tasks. Donders utilized three tasks:

Task a = one possible response to one possible stimulus (simple RT).

Task b = two stimuli, two responses with a one to one mapping between them (the most common choice RT paradigm).

Task c = two stimuli and only a single response required for one stimulus, but not the other.

Task a was presumed to involve only the process of simple response. Task b presumably involves three processes: stimulus categorization, response selection, and simple response. Task c presumably involves two processes: stimulus categorization and simple response. Reaction times for each task fell into the expected rank order: RTb > RTc > RTa. If RTb is

longer than RTc, and RTc is longer than RTa because of the sequential addition of another "mental process," then the reaction times associated with each process can be indirectly obtained via subtraction. That is, stimulus categorization = RTc - RTa, and response selection = RTb - RTc. Thus, a particular choice reaction time paradigm could be developed specifically for the purpose of evaluating such processes as response selection and stimulus categorization. This remains unchanged, though the subtraction method has been replaced by more sophisticated procedures.

Currently, two of the most widely cited and supported theories of information processing are processing stage theory (Sternberg, 1969b) and multiple resource theory (Wickens, 1984). Both of these theories assert that humans possess several different capacities with resource properties. These theories and related studies (Wickens, 1984) generally posit three primary stages of processing and associated resources dedicated to perceptual input, central processing, and motor output. A choice reaction time paradigm can systematically influence any of these three stages. For example, varying the perceptibility and/or modality of the stimulus could influence stimulus encoding (input processing), the stimulus response mapping ratio and/or compatibility (sameness of spatial orientation) can influence stimulus categorization and response selection (central processing), and prescribed response activity as well as modality of response can affect the motor output processing stage. Thus, choice reaction time paradigms can be very useful with reference to such an information processing framework. Via systematic manipulation of stimuli, stimulus response compatibility mappings, and responses within an experiment, the relationships (e.g., independence versus parallelism) among processing resources can be evaluated by examining the resulting statistical relationships among the reaction times obtained under the different task conditions (Sternberg, 1969b). By the same token, a multiple resources information processing framework can be very useful in the design of a choice reaction time task. Based on this framework, a choice reaction time task can be designed to primarily tap resources dedicated to a given processing stage, although the choice reaction time paradigm necessarily involves resources from all processing stages to at least some degree. Thus, a choice reaction time

90

task can serve well as a diagnostic tool in the assessment of the influence of environmental stressors on particular processing resources.

The UTC-PAB version of the choice reaction time paradigm, as has been described, is a four-choice task. There are four possible stimuli variations, each associated with one correct response key on a keypad. Thus, there is a 1:1 mapping of stimulus onto response. Also, the stimulus response compatibility associated with the task is very high. These two task characteristics, when considered in an information processing framework (Sternberg, 1969b; Wickens, 1984), necessitate the assignment of the task demands primarily to resources dedicated to perceptual encoding. Neither this mapping ratio nor this level of compatibility is associated with heavy demands on central processing resources (stimulus categorization or response selection) or motor output resources (Smith, 1968).

When a 1:1 mapping of stimulus onto response is used, mean choice reaction time has been found to increase linearly with LOG_2 of the number of alternative stimuli. This finding is readily explained utilizing information theory which defines a bit of information as LOG_2 of the number of possible alternatives (Hick, 1952). That is, mean choice reaction time increases linearly with bits of information, indicating an increase in processing demand with a greater number of alternative stimuli. Most likely, this demand is primarily placed on perceptual input processing resources, though obviously some stimulus categorization is necessary. The only stimulus characteristic being varied is its location, so a high level of categorization is not required. Studies explicitly designed to study the stimulus categorization process frequently employ many 1:1 mappings, the response being required if a stimulus is representative of a particular category (Smith, 1968). For example, a subject may be told to respond if the stimulus which apears is a member of a particular set (i.e., if it is a vowel). This involves a high level of categorization, requiring the activation of memorial resources which are unquestionably part of the central processing stage.

While the stimulus response mapping seems to limit the demand on central processing, the high degree of stimulus response compatibility associated

with the UTC-PAB version of the four-choice reaction time task would seem limited to the demands placed on central processing and motor output resources. Studies which attempt to directly evaluate the response selection process often manipulate the stimulus response compatibility, requiring the subject to mentally perform a spatial reorientaiton of the stimulus display or the response apparatus to reduce any incompatibility and deliver the appropriate response. The mental manipulation of spatial information is also usually considered a central processing resource (Wickens, 1984).

In the UTC-PAB version of the task, the stimulus display and response apparatus are formatted in a fashion which has virtuaily no inherent incompatibility, and experimenter instructions never change this. Motor output resources are also frequently thought to be involved in the process of response selection (Shingledecker, 1984) which, as mentioned, plays a limited role in the UTC-PAB version of the four-choice reaction time task.

In summary, the UTC-PAB four-choice reaction time task has several built-in advantages. The potential for portability and self administration of the original task (Wilkinson and Houghton, 1975; Ryman, Naitoh, and Englund, 1984) has led to the employment of this task in many studies of environmental stressors. Thus, its reliability and sensitivity have been documented (see sections on reliability and sensitivity). Also, the fact that this task represents a variation of a traditional paradigm allows for the interpretation of task sensitivity within an information processing framework. The stimulus and response characteristics of this version of the task would seem to place demands primarily on perceptual input resources, though any choice reaction time task necessarily places at least minimal demands on all three stages of processing (Smith, 1968). It should be noted, however, that many studies that have utilized this task in the investigation of stressors have not been concerned with the ramifications of information processing theory, and results are not interpreted in these terms, although the potential for such interpretation is always present when utilizing a choice reaction time paradigm.

RELIABILITY

According to Krause and Bittner (1982), the four-choice RT task appears to be characterized by sufficient internal reliability. In arriving at this conclusion, Krause and Bittner computed intersession correlations for three performance measures: reaction time (RT), movement time (MT), and total time (TT; see Table 10 for Response Measures). These correlations were performed on data obtained using one-, two-, and four-choice RT tasks. It was determined that general measures associated with one- and four-choice tasks were generally stable, especially RT and TT. The actual correlation values associated with the four-choice task were as follows: RT = .68, MT = .86, and TT = .82. There were 15 subjects, all Navy enlisted men. Fifty trials on each of the three conditions (one-, two-, and four-choice) were presented in blocks. Each subject completed 1 block per day for 15 consecutive workdays. Therefore, each subject was presented with 2250 trials, 750 at each condition; and subjects were never confronted with more than one condition on any given day. Krause and Bittner also performed stability analyses on this data for all conditions and measures. For the four-choice RT task, MT values were found to stabilize on day nine, TT values on day 10, and RT values on day 11 (note: differential stability is characterized by high, stable test-retest correlations). Based on these findings, Krause and Bittner conclude that "four-choice RT measures are generally stable and are recommended for inclusion in performance assessment batteries, with at least 1000 practice trials prior to repeated measures applications" (p. 5). It can then be inferred that the UTC-PAB four-choice RT task is sufficiently reliable and stable for performance assessment applications as it is a four-choice RT task and the principal performance measure is analogous to the TT measure investigated by Krause and Bittner (1982).

VALIDITY

In their development of a portable four-choice RT paradigm, Wilkinson and Houghton (1975) attempted to establish a preliminary framework of performance norms for the task. Three performance measures were obtained: reaction time, mean number of gaps, and mean number of errors (see Table 10 for

Response Measures). The subjects were five enlisted men who were required to perform 20 minutes of continuous responding following 5 minutes of practice and a 5-minute break. Mean values for each of the three performance measures were calculated for each of the four 5-minute segments of the total 20 minutes. Mean scores were also obtained for the initial five 1-minute periods. The data were then examined at two levels: (1) comparisons among scores within each set of scores, and (2) overall comparisons between the two sets of scores. The results showed that all three performance measures decreased as a function of elapsed time on task. This effect of fatigue was in accordance with the expectations of Wilkinson and Houghton (as per the five-choice serial RT task of Leonard, 1959, from which much of the procedural framework of the four-choice task is borrowed). Of particular interest to the issue of task validity are the correlations that were calculated among the three performance measures. That is, scores on each performance measure were compared with scores on the two remaining measures. The results were as follows:

RT versus GAPS, $r = +.90$ (p < .02)
RT versus ERRORS, $r = +.83$ (p < .05)
GAPS versus ERRORS, $r = +.88$ (p < .025)

Also, Kendall's concordance measure across individuals among the three within test deterioration scores (the difference between first half and second half scores) was .844 (p < .01). Thus, all three scores agreed with each other in reflecting an overall deterioration in performance during the test.

In conclusion, based on the data obtained and the analyses performed by Wilkinson and Houghton (1975), it can be stated that the four-choice RT task appears to be characterized by considerable internal validity. That is, potential task sensitivity to a stressor (fatigue in this case) is probably not heavily dependent upon the particular performance measure or individual subject being evaluated. Performance decrements associated with the four-choice RT task can be attributed with a reasonable degree of certainty to the experimental manipulations being evaluated, as such decrements are likely not limited to the measures or subjects involved.

94

Most studies which typically employ four-choice RT tasks as a diagnostic tool can be divided into two general categories; those which attempt to evaluate the effects of a particular drug or drugs and those which attempt to evaluate the effects of fatigue, either due to physical effort, sleep loss, or both.

Cherry et al. (1983) investigated the potential influence of toluene and alcohol on psychomotor performance. Four-choice RT was one of four diagnostic tests utilized to assess psychomotor performance. The authors' interest in these two drugs was due to the roles these two chemicals can play in certain occupational environments. Toluene is a benzene analogue which can be used as a rubber solvent, in paints and varnishes, in printing, and in glues and adhesives. It is possible that occupational exposure to toluene and the consumption of alcohol may separately, or in combination, impair psychomotor performance, diminishing operator productivity and/or safety. Mean blood levels for alcohol and toluene were 49.9 mg percent and 12.7 mmol/1 respectively. Surprisingly, neither drug exerted a significant main effect on mean reaction times obtained on the 4-choice task. The alcohol X toluene interaction was also nonsignificant. Perhaps these results were partially due to the great degree of intersubject variability present in this study. In fact, when subjects are entered into the analysis as a random source of variation, the resulting F value was significant ($F = 72.2$, $p < .001$). Also significant in this analysis were the alcohol X subjects ($F = 18.1$, $p < .01$), the toluene X subjects ($F = 27.0$, $p < .001$), and the alcohol X toluene X subjects ($F = 4.2$, $p < .05$) interactions. Thus, it appears that the potential effects of these drugs on four-choice reaction time performance are largely a function of the subject(s) involved. In other words, these two drugs produced performance decrements for some subjects, but did not affect the performance of others. The salience of the subject variation in the analysis could be due to the employment of a rather small subject pool ($N = 8$).

95

A four-choice serial reaction time task was employed by Herbert, Mealy, Bourke, Fletcher, and Rose (1983) to assess the effects of general anaesthesia on the individual's recovery of mental functioning. The prescribed task parameters and apparatus were precise as per Wilkinson and Houghton (1975; that is, a portable cassette recorder was appropriately modified, and data were stored on magnetic tape). Each of the 10 test blocks lasted 5 minutes. The 55 subjects were divided into four experimental groups, based upon varying method of anaesthesia and modes of ventilation in recovery. The four groups can be labeled as follows:

(1) Halothane (anaesthesia), spontaneous ventilation.
(2) Standard anaesthesia (thopentone 250 mg, halothane, 0.5 - 1.5 percent, nitrous oxide, and oxygen), spontaneous ventilation.
(3) Standard anaesthesia, controlled ventilation.
(4) Control (12 orthopedic hospital patients who had not had an operation for at least two weeks).

All experimental groups (one, two, and three) showed significant impairment (with reference to the control group) on the four-choice RT task 90 minutes after regaining consciousness. Significant impairment remained on post-operative day one for only group one (p < .05). This being the case, the findings on post operative day two were somewhat surprising. On day two, the mean RTs of groups one and two were again significantly different from those of the control group. It should be noted that this was largely due to the improvement in performance of the control group; possibly a practice effect. Group three also improved, while groups one and two did not markedly improve from day one to day two. Perhaps controlled ventilation, from a mental processing frame of reference, enhances one's recovery from anaesthesia.

These findings on post operative day two bring to light the importance of a reliable diagnostic test of psychomotor performance following exposure to anaesthetic drugs, as recovery would be expected by this time. This point is reinforced by the subjective ratings obtained by Herbert et al. (1983). On day two, group three subjects felt subjective impairment to a greater extent than did groups one and two. Thus, the subject who reports

96

a return of energy and alertness may not necessarily be able to perform in accordance with such reports.

Englund, Naitoh, and Ryman have utilized their computer administered version of the four-choice reaction time paradigm (Ryman, Naitoh, and Englund, 1984) in recent investigations on the effects of sustained physical effort (Englund, Naitoh, and Ryman, 1984; Englund et al., 1985). The subjects involved in these studies were physically fit male marines (N = 40), and the physical effort for the experimental group consisted of walking on a treadmill while wearing full combat gear and packing a rifle for the first 30 minutes of each one hour session. The control group subjects were not subjected to these conditions for the first 30 minutes of each one hour session. In the second 30 minutes of each session, all subjects were required to perform a number of cognitive tasks, including four-choice reaction time. There were no significant group differences in either mean reaction time or percent correct. The only significant effect associated with the four-choice task was a time of day effect on percent correct. Accuracy was significantly lower (79.5 percent) during the last session (session 17) than it was for all previous administrations (85.2 percent - 87.7 percent). Repetition of the task may have been fatiguing, but the required physical effort of the experimental group was not fatiguing with reference to the four-choice reaction time task.

The Wilkinson and Houghton (1975) version of this task has been frequently utilized in studies concerned with sleep deprivation effects (Angus and Heslegrave, 1985; Bonnet, 1980; Glenville et al., 1978; Glenville and Wilkinson, 1979; Taub, 1982; Tilley et al., 1982). The specific findings of these studies with reference to the four-choice reaction time task are highly consistent. Extended periods without sleep were seen to produce significant decrements in mean reaction time, while accuracy levels (i.e., percent correct) remained unaffected. In addition, Glenville and Wilkinson (1979) noted an increase in the number of gaps (see Table 10 for Response Measures) for sleep deprived subjects. Also noted in the sleep loss literature were significant decrements in mean reaction time associated with time on task. Time on task reaction time decrements were previously associated with the development of the task (Wilkinson and Houghton, 1975), and

time on task was also the cause of decreased performance accuracy in the investigations of physical effort (Englund, Naitoh, and Ryman, 1984; Englund et al., 1985). The potential of this task to show performance decrements simply as a result of task repetition should be taken into consideration whenever this task is employed.

Four-choice reaction time tasks have been employed in dual task paradigms which are designed to test particular aspects of the previously discussed information processing framework (Kantowitz, Hart, and Bortolussi, 1983; Looper, 1976). In each of these studies, four-choice tasks are performed in conjunction with tracking tasks. The goal is to assess difficulty of tracking conditions via performance on the four-choice task. The results indicate that the four-choice task performance is a reliable indicator of tracking difficulty. Reaction times consistently increase as tracking difficulty is increased. This finding can be accounted for within the previously discussed framework of information processing. The four-choice reaction time task primarily taps perceptual encoding resources which are also necessarily engaged in a tracking task. Performance decrements associated with this dual task combination are probably due to the heavy demands being placed on these resources. Four-choice reaction time tasks are, thus, useful in the investigation of information processing resources because they are often sensitive to dual task conditions.

TECHNICAL DESCRIPTION

The experimenter initializes the task and instructions appear on the screen. The actual task begins after the subject makes the first key press response. The screen is then divided into four quadrants. A cursor with the blinking plus sign appears in one of the quadrants. The blinking plus sign is sent to a randomly selected quadrant following a response. The program performs a random select from the response time of the subject in the following way: the last reaction time (last two bits) is divided by four. If the remainder is zero, then the cursor is sent to the upper left quadrant. If the remainder is one, then the quadrant selected is the upper right; if two, lower left; if three, lower right. An auditory signal is

98

sounded after 2.5 seconds if there has been no response and continues until a response is made.

Trial Specifications

The "+" will remain in a particular quadrant until the subject presses a response key. Immediately following the response, the quadrants will blank and will remain blank until the next trial when the "+" will reappear in one of the quadrants, and the subject is again required to press the appropriate key. It is recommended that trials be separated by a brief (about one second), constant interstimulus interval (ISI). If the subject responds during the ISI, an "error message" should appear on the screen (e.g., "please do not respond until the '+' appears").

DATA SPECIFICATIONS

Reaction times of all responses are recorded in milliseconds. Incorrect (wrong quadrant) responses and lapses (gaps) of 2.5 seconds are also tabulated.

The following summary statistics for reaction times are provided: the mean and standard deviation of all correct responses, incorrect responses, the 10 percent fastest correct responses, the 10 percent slowest correct responses, the 10 percent fastest incorrect responses, and the 10 percent slowest incorrect responses. A percent correct response value is also provided (see Table 10 for Response Measures for other measures which have been used with this task).

TRAINING REQUIREMENTS

Subjects are told that this task is a test of their reaction time and their ability to choose the correct one of four choices. Following the presentation of the instructions, subjects should perform two 6-minute blocks of training trials. The experimenter should carefully evaluate training performance to insure that instructions are being followed. The most

important aspect of the instructions to be emphasized is that subjects are to try to respond as quickly and accurately as possible.

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

There are minimal training requirements for this task. Subjects usually reach proficiency in one or two 6-minute practice blocks.

INSTRUCTIONS TO SUBJECTS

A blinking "plus sign" will be presented in one of the four quadrants of the CRT. The object of the four-choice serial reaction time task is to press the key on the keyboard that corresponds to the quadrant with the blinking plus sign. The blinking plus sign remains in a given quadrant until one of the four keys is pressed and then randomly appears in any one of the four quadrants, at which time you again press the corresponding key on the keyboard. This process continues for 6 minutes. Respond as quickly and accurately as possible. If none of the four keys is pressed within 2.5 seconds of the onset of the blinking plus sign, a bell rings every .1 seconds until a response is made. Reaction times of all responses, correct and incorrect, are recorded. Press any of the four keys to start the sequence.

Section 9
ALPHA-NUMERIC VISUAL VIGILANCE TASK (UTC-PAB TEST NO. 8)
(SUSTAINED VISUAL ATTENTION--CHOICE RT)

PURPOSE

The purpose of the Alpha-Numeric Visual Vigilance Task (ANVVT) is to test a
subject's ability to continue making decisions and rapid responses to vis-
ual symbols for long nonstop periods. The ANVVT is a discrimination reac-
tion task intended to simulate a situation in which a person monitoring a
visual display might show fatigue and performance decrement without being
aware of it.

DESCRIPTION

The UTC-PAB version of the ANVVT consists of CRT presentation of random
alphabetic characters or numbers at random intervals ranging between 6 and
14 seconds, with a mean interval of 10 seconds. The number or character,
10 by 28 mm in size, remains on the screen for 500 msec. Subjects are
instructed to press a hand held, normally open push button switch with
their thumb every time an "A" or a "3" appears. No response is required to
other stimuli.

Twenty "As" and "3s" are randomly mixed with 160 other characters and num-
bers given during this 30-minute task. Response latencies and errors are
recorded. There are two types of possible errors: (1) errors of commis-
sion (responding to non "As" and non "3s"), and (2) errors of omission (not
responding to an "A" or "3" in 5 seconds). Reaction times are recorded in
msec for both correct responses and errors of commission. Errors of omis-
sion are scored as reaction times of 5000 msec.

BACKGROUND

The vigilance task has been regarded as providing "the fundamental paradigm
for defining sustained attention as a behavioral category" (Jerison,
1977). Research on the topic of sustained attention or vigilance is

concerned with the ability of observers to detect signals over prolonged periods of time. The theoretical importance of the vigilance situation is that it allows one to study in a simple and controlled task almost all of the factors which may be considered to influence attention.

The ANVVT (Hord, 1982; Naitoh, 1981) was developed at the Naval Health Research Center to measure long term visual vigilance. The ANVVT was adapted from the Continuous Performance Task (CPT) devised by Rosvold et al. (1956) to study brain damage. The CPT is a cognitive vigilance task which consists of the presentation of a series of letters in which each occurrence either of one letter (e.g., A) or of a sequence of two letters (e.g., AX) has to be detected. Letter stimuli are generally presented at a rate of one per second for a 10-minute period. Target letter(s) occur irregularly throughout the series and represent 25 percent of all stimulus presentations. The CPT can be presented both visually and auditorially. Only positive stimuli are responded to and a response deadline is set at 0.7 seconds. Three possible types of errors include responses with a latency longer than 0.7 seconds (late correct responses), failure to respond to the stimuli (errors of omission), and responses to other stimuli (errors of commission).

The ANVVT differs from the CPT in that numeric characters as well as alphabetic characters are presented as stimuli. Also, subjects do not monitor the occurrence of two character sequences. The task is of longer duration (e.g., 30 minutes), however, stimuli occur less frequently (mean interval of one per every 10 seconds).

In all instances in the literature, the CPT and ANVVT tasks have been utilized with variables known to effect attention processes in order to determine if an attention deficit is obtained. As mentioned earlier, the CPT was originally developed as a diagnostic instrument for the investigation of brain damage. Brain damaged patients make generally more errors on this task than do normals, and the difference in error rates increases in the more difficult A-X version, in which a greater memory load is imposed (Rosvold et al., 1956). Also, the brain damage impairment is likely to reveal itself in the form of attentional lapses rather than as a

steady decline of detection efficiency. Alexander (1973) also used the CPT in a comparison of the performance of hospital patients with either organic senile dementia, or patients in whom brain damage had not been diagnosed, and of a group of nonhospitalized subjects. He found that the senile dementia group detected significantly fewer signals than did either of the control groups and that this group was also the only one to make more false alarms (errors of commission) than errors of omissions.

Other experiments have demonstrated that older subjects who have not sustained brain injury also perform worse on the CPT than do younger subjects (Canestrari, 1962; Davies and Davies, 1975). In these versions of the CPT, responses made within 700 msec following a signal are scored as correct detections, while responses made after this period has elapsed are scored as errors. Thus, performance on the CPT may not reflect solely a change in the capacity to sustain attention, but may also be a consequence of the well established loss of response speed that accompanies normal aging and which also results from brain injury. Davies and Davies (1975) analyzed their CPT data in detail and attempted to separate false alarms from other errors. They found no age differences in false alarm rates but did obtain a highly reliable effect of age for errors which includes slow correct detections. Older men, between the ages of 63 and 72 years, made many more of these errors than did younger men between the ages of 18 and 31 years.

The CPT has also been used to determine the effects of temperament and hyperactivity on sustained attention. Hogan (1966) found that introverts detected significantly more signals than did extroverts on a 10 minute visual version of the CPT. Sykes, Douglas, and Morgenstern (1973) compared the performance of hyperactive children to normal children on the CPT. An impairment in performance was found; hyperactive children detected fewer signals and made more overall incorrect responses than normal children. In addition, while the performance of hyperactives declined with time on task on the 15 minute CPT, no decrement was observed for normal children.

In an experiment by Mirsky and Cardon (1962), attentive behavior (measured by the CPT) and EEG were studied simultaneously in normal subjects under the influence of sleep loss or the depressant drug chlorpromazine. Both

sleep deprivation of 66 hours and administration of 200 mg chlorpromazine
were found to significantly decrease performance on the CPT. An analyses
of errors showed that late correct responses occurred, on the average to
fewer than five percent of the positive stimuli, whereas errors of omission
occurred, on the average, to almost 24 percent of the positive stimuli.
Errors of commission occurred infrequently in all conditions. EEG analysis
indicated slow wave changes during error periods of performance on the CPT
for sleep deprived subjects, but not for subjects receiving chlorproma-
zine. The significance of these findings was discussed in relation to the
possible existence of separate, but closely related mechanisms within the
reticular activating system, which mediates behavior on the one hand and
the EEG on the other. That is, the two groups (sleep deprived and drug
groups) were similar in terms of performance, but differed with respect to
their EEG patterns.

The earliest use of the ANVVT was an experiment conducted by Townsend and
Johnson (1979) that also examined the relation of EEG to sustained atten-
tion with sleep deprived subjects. A 3-hour version of the ANVVT was per-
formed on four consecutive days, with the task on the third day preceded by
one night of total sleep loss to maximize drowsiness and associated per-
formance decrement. If the alpha-numeric character was an "A" or "3"
(34 occurrences/h), the subject responded by pressing one switch; if the
character was any other letter or number (326 occurrences/h) the subject
responded by pressing a second response switch. Reaction time in msec, as
well as EEG from stimulus onset to subjects' response was recorded. The
analysis was conducted on the 10 shortest and 10 longest RTs, and 10 trials
where the subject failed to respond. Significant univariate correlations
were found between RT and the frequencies in the 15 to 20 Hz range of EEG
activity. A multiple regression analysis using up to five EEG frequencies
indicated significant correlations of prestimulus EEG activity with RT.
The results suggest that sleep deprivation did increase the contribution of
drowsiness related EEG change and, thus, improved the EEG-RT correlation.

Hord (1982), in a related study, examined the relationship between EEG and
reaction time within subjects, such that the EEG could be used to predict

perfonnance decrement in vigilance situations before the decrement
occurs. Subjects who had not been previously sleep deprived performed the
ANVVT on three consecutive days. These results showed no major changes in
mean reaction time and errors of omission during the 3-hour test period.
The ratio of the sum of intensities in the 1 to 6 Hz to 7 to 12 Hz band was
obtained for each condition (10 fastest trials, 10 slowest trials, errors
of omission). The group mean ratios for the three conditions indicate
little difference between fast and slow trials, but a big difference
between errors of omission and the other two (fast and slow). The author
concluded that: (1) EEG predictors of perfcrmance change during monitoring
can work in situations where the subjects had not been previously sleep
deprived. (2) The predictive power of the EEG ratio may not be practical
by the third day because of the increased error of omission rate during the
middle of the session. (3) The EEG ratio is certainly simpler to implement
than the stepwise multiple regression approach as used by Townsend and
Johnson (1979).

In summary, the ANVVT is an adaptation of the continuous performance
task. These cognitive vigilance tasks are short duration tests of sus-
tained attention perfonmance. The CPT has been used to examine conditions
which are known to effect attention processes (e.g., brain damage, age,
sleep loss, and drugs). The ANVVT has primarily been used to determine if
there are any physiological correlates (e.g., EEG) of performance decre-
ments on vigilance tasks.

RELIABILITY

No studies have been conducted that directly assess the reliability of the
ANVVT. Thus, there is little indication that repeated performance on this
task will produce similar results. Some reliability information may, how-
ever, be inferred from the results obtained by Hord (1982). Subjects in
this study perfonmed the ANVVT for 3-hour periods on each of three con-
secutive days. Results showed no major changes in reaction time and errors
of omission during the 3-hour test period for each day. It also appeared
that mean reaction times declined over the three days while mean errors of
omission tended to increase. Thus, it appeared that performance scores on

this task remained relatively stable over repeated testing periods. However, until actual performance intercorrelations are reported, the true reliability of this task remains uncertain.

VALIDITY

The ANVVT has been used to measure sustained attention performance. More specifically, the task attempts to test a subject's ability to continue making visual detections and discriminations over a period of time. The task is closely related to the continuous performance task which is a well known and more established cognitive vigilance task. The CPT has been shown to reflect attentional decrements in many studies and with a variety of variables known to be sensitive to sustained attention (e.g., age, temperament, hyperactivity, sleep loss, and drugs). The ANVVT has not yet established the degree of validity set by the CPT, but the two tasks do seem to measure the same mechanisms of attention. Experiments using the ANVVT with a greater variety of variables and obtaining significant results would greatly increase the validity of the task as a measure of sustained attention.

SENSITIVITY

Experiments demonstrating the sensitivity of the continuous performance task to a number of attention related variables have already been discussed. It was also stated in the background section of this manual that the ANVVT was found to be sensitive to sleep loss in a study relating EEG to reaction time (Townsend and Johnson, 1979).

Other uses of the ANVVT have utilized the task as a measure of cognitive vigilance performance during sustained operation episodes (Englund et al., 1983; Englund et al., 1985). In both of these studies, the effects of physical work, sleep loss, continuous work (CW), and time of day on various cognitive and physiological tasks were assessed. All subjects performed every task on each of three consecutive days. Day two and day three represented the two continuous work episodes (CW1, CW2) and were separated by a 3-hour nap midway between sustained episodes. Physical work was

106

manipulated by having half the subjects perform the ANVVT while walking a treadmill (at 30 percent of VO_2 maximum), the other subjects performed the ANVVT while seated in front of a CRT.

The ANVVT was given the first half hour of each 1 hour session during both CW1 and CW2; thus, subjects completed this task 17 times per CW episode. In the task, random alphabetical or numerical characters were presented on the screen at random intervals between 6 to 14 seconds (mean interval of 10 seconds). The numbers remained on the screen for 10 msec. Subjects were instructed to press a button every time an "A" or a "3" appeared. The task lasted for 30 minutes, during which 20 signal stimuli were randomly mixed with 100 other characters. Percent of correct responses was used as the dependent measure.

Results from physiological measurements, such as oral temperature, heart rate, blood pressure, and grip strength are reported in Englund et al. (1983). Cognitive test results (e.g., logical reasoning, air defense game, and four choice RT) are reported in Englund et al. (1985). Both studies report results for the ANVVT. Analysis of the ANVVT data indicated a significant interaction involving groups. The exercise group improved in performance during CW1, whereas the control group's performance was essentially the same across the first day. During CW2, the exercise group showed the same slight improvement during the first half of the day as in CW1, and then significantly declined in percent detections during the second half of CW2. The control group indicated significantly lower performance during CW2. Performance on the ANVVT also indicated a significant day difference. The mean percent correct detections was 80.9 percent during CW1, but only 70.6 percent during CW2. Mean errors of omission increased by 55 percent from CW1 to CW2 and mean reaction times increased by 25 percent. The results from this study indicated that moderate exercise does not combine with sleep loss to further decrease cognitive performance.

TECHNICAL DESCRIPTION

Twenty "As" and "3s" are randomly mixed with 16 other characters and numbers. The stimuli are selected from a list of numbers and letters

107

randomized every run. This list is stored within the program. The random
intervals for alphabetic character/number presentations range between 6 and
14 seconds, with a mean interval of 10 seconds. The number or character is
10 by 20 mm in size and remains on the screen for 500 msec. The task lasts
for 30 minutes at which time an auditory signal is sounded. The program
measures response latencies. At the end of a 30-minute task, all reaction
times in milliseconds are stored. Errors of omissions (no response to an
"A" or a "3" in 5 seconds) are stored as 5000 msec latencies.

DATA SPECIFICATIONS

The listing scoring program for the alphanumeric task lists all responses
during a 30-minute session, the number of correct responses (button presses
following an "A" or "3"), the number of errors of omission, and the number
of errors of commission. The means and standard deviations for the correct
responses, the five slowest correct responses, and the five fastest correct
responses are also printed out, along with the percent correct responses
and percent correct detections. An error of omission is declared when
responses to an "A" or a "3" are not made within 5 seconds. In computing
mean reaction times as well as the five slowest responses, errors of omis-
sion are added as reaction times of 5000 msec (5 seconds).

TRAINING REQUIREMENTS

The instructions should be read to the subject before the start of the
training trials. Extensive practice is not required for this task. One or
two sets are usually sufficient to familiarize the subject with the char-
acteristics of the task and target stimuli.

To summarize, the training phase for this test should consist of the
following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that
    the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

In this experiment, you are to monitor the TV screen on which alphabetic or numerical characters will be briefly flashed. One randomly selected alpha-numeric character will be presented every 6 to 14 seconds. If the character is an "A" or a "3" you are to respond by pressing the designated switch. If the character is any letter/number other than "A" or "3" no response is required. Please respond as quickly and accurately as possible. The task will last for 30 minutes.

Section 10
MEMORY SEARCH TASKS (UTC-PAB TEST NO. 9)
(SHORT TERM WORKING MEMORY--AUDITORY AND VISUAL MODALITIES)

PURPOSE

The purpose of this memory search task is to test a subject's ability to make comparisons of letters maintained in memory. The task is diagnostic of the processes of selective retrieval and comparison in short term work-ing memory. This task may also reflect processes involved in the encoding of stimulus items, categorization, response selection, and response execution.

DESCRIPTION

Either one, two, four, or six alphabetic characters make up the "positive set" which is presented to the subject to maintain in memory. The remain-ing alphabetic characters make up the "negative set." Subsequent to the presentation of the "positive set," individual probe letters are presented to the subject for comparison and classification as being members of the positive set or the negative set. Subjects respond by pressing the appro-priate key on a two button keypad.

There are three different procedures used in this task. Each procedure is presented in a visual version and an auditory version making a total of six unique versions. In the varied set procedure (VS) a different positive set is generated on every trial followed by a single probe item. The fixed set procedure (FS) involves the presentation of the positive set followed by 100 probes to constitute a trial. A trial in the mixed set procedure (MS) consists of the presentation of 10 separate positive sets of equivalent size, each of which is followed by 10 probes for classification with respect to the set. In the visual versions (V) of these procedures, all stimuli are presented on a CRT, and in the auditory versions (A) the probe items are presented via a speech synthesis system and positive sets are presented both visually and auditorially.

110

BACKGROUND

The use of results from reaction time (RT) experiments to study stages of information: processing began about a century ago with a paper titled, "On the Speed of Mental Processes," by F. C. Donders (1969). In the paper Donders introduced the "subtraction method" (a method for analyzing the RT into its components and thereby studying the corresponding stages of processing). To use the subtraction method one constructs two different tasks in which RT can be measured, where the second task is thought to require all the mental operations of the first, plus an additional inserted operation. The difference between mean RTs in the two tasks is interpreted as an estimate of the duration of the inserted stage. This interpretation depends on an assumption of pure insertion which states that changing from task one to task two merely inserts a new processing stage without altering the others.

After a brief popularity, this technique fell out of favor. It was found that the elements of cognitive performance were not independent, and that they could not be treated by a simple additive, linear model. This criticism was insurmountable with the mathematical techniques available at the time and efforts to probe cognition diminished for a long time.

With proper statistical control, independence of stages can presently be determined (Sternberg, 1969b). Modern experimental methodology and data analysis led to applications of the stage theory that seem to withstand the early criticisms. One such application provided by Sternberg (1969a) focused on mechanisms of memory retrieval for information in both short term and long term memory. The approach is also being widely used to confront issues such as what information is stored and how it is coded and organized. Sternberg (1969a) used individual symbols as units to be remembered, and gained control over the "memory load" under which the subject was operating. The desire to analyze the processing of information into its functional components (particularly when combined with the hypothesis that component processes are arranged in stages) leads naturally to RT methods and to an interest in the temporal parameters of processing.

111

The purpose of the memory search tasks is to study the ways in which information is retrieved from memory when learning and retention are essentially perfect. The method involves the presentation of a list of items (e.g., letters) for memorization that is short enough to be within a persons immediate (short term) memory span. The subject is then asked a simple question about the memorized list to which a quick response is made, and the delay in responding is measured. By examining the pattern of this RT, while varying such factors as the number of items in the list and the kind of question asked, one can make inferences about the underlying retrieval processes.

The remainder of this section will describe the various factors which affect memory scanning processes. Various models and procedures as well as their predictions will be outlined. Finally, some of the extensions and generalizations of the early findings of memory scanning tasks will be presented.

The Item Recognition Paradigm

The Item Recognition Paradigm is a particular experiment designed by Sternberg (1969a) which allows control over the short term memory load of a subject. In the paradigm, the "stimulus ensemble" consists of all the items that might appear as test stimuli (e.g., the letters of the alphabet, the numbers 0 to 9). From the ensemble a set of elements is selected arbitrarily and is defined as the positive set. (The positive set size selected is usually an independent variable in the experiment. Sizes may vary from one to nine elements but should not exceed the subject's short term memory capacity). The items comprising the positive set are presented as a list for the subject to memorize. The remaining items in the ensemble are called the negative set. When a test stimulus or "probe item" (one item randomly chosen from the stimulus ensemble) is presented, the subject must make a decision as to the appropriate membership of that item. If the probe item is a member of the positive set, the subject presses a predetermined button. If the item is a member of the negative set, an alternate button is pressed. The RT is measured from the onset of the test stimulus

112

to the response. It is a requirement of the procedure that virtually error free performance is maintained (error rate < 2 percent).

Within the Item Recognition Paradigm, different procedures can be used. In the varied set procedure, the subject must memorize a new positive set on each trial. The set may be presented all at once (in parallel) or serially, followed by a retention interval of 2 or 3 seconds during which the subject is free to rehearse, then a warning signal, and then a test stimulus. In the fixed set procedure, the same positive set is used for a long series of trials, and a trial consists only of warning signal, test stimulus, and response. In the varied set procedure, positive set items are stored and rehearsed in short term memory only. Whereas in the fixed set procedure, positive set items are believed to be stored in the long term store. However, the similarity òf results from the two procedures suggests that the same memory system was being scanned. That is, when information in long term memory has to be used, it may be transferred into short term memory (where it is maintained by rehearsal) and, thus, becomes more readily available.

Set Size Effects

The main variable investigated in memory scanning studies is the effect of the size of the positive set on the response time, while keeping constant the relative frequency with which positive and negative responses are required. If the average reaction time is plotted as a function of the memory set size, then the resulting function represents the subject's ability to make memory based decisions. Four features of this function should be noted (Figure 7): (a) mean RT increases approximately linearly with set size; (b) the rate of increase is the same for positive and negative responses; (c) the rate of increase is about 38 msec for each item in the positive set; and (d) the zero intercept is about 400 msec. It can be seen that the slope of the function generated in a Sternberg task represents the internal "processing efficiency" of the short term memory system. This function is obtained regardless of the procedure used, varied or fixed set. The remarkable similarity of results from the two procedures
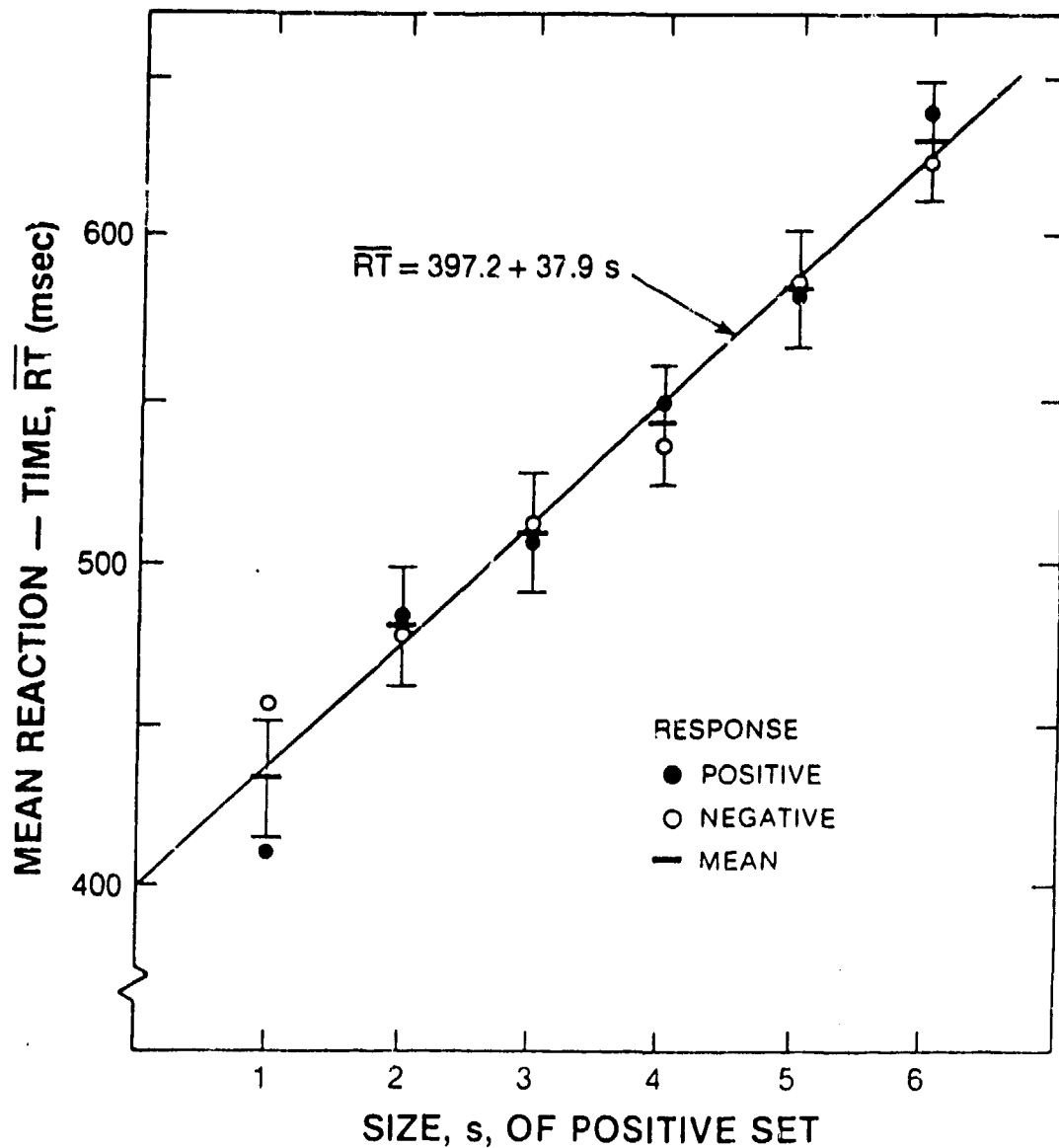
113

Figure 7. Results of Experiment 1 from Sternberg (1969a) Which
Utilized the Varied Set Procedure [Mean Latencies of
Correct Positive and Negative Responses, and Their Mean,
as a Function of Size of Positive Set. Averaged Data
From Eight Subjects, with Estimates of ±5 About Means,
and Line Fitted by Least Squares to Means]

indicates that the same retrieval process was used for both the unfamiliar and well learned lists.

The size of the negative set has also been varied in this paradigm while maintaining a constant positive set size. Here, mean RT is plotted as a function of the size of the negative set. The size of the negative set had no significant effect on the overall mean RT. This implies that the ensemble size per se has no effect on memory scanning times.

Models and Predictions

How does a person decide whether the test stimulus is contained in the positive set? That is, in what manner is the test stimulus compared to the items of the positive set which exist in memory. Several models of this memory search process have been proposed. Each model leads to a different prediction of search functions which can be verified through experiments utilizing the item recognition paradigm.

One possible model to describe the processes of memory search is a parallel comparison model. In this model, the test stimulus is compared in parallel to all members of the positive set. The particular parallel model that has attracted most attention has been considered by Atkinson, Holmgren, and Juolea (1969) and Townsend (1971). According to this model all comparisons start simultaneously and have durations that are exponentially distributed. Each of the simultaneous comparisons is assumed to require processing capacity. There is a fixed amount of resources which is equally divided among those comparisons not yet completed. The increase of mean RT with set size is assumed to result from the sharing of the fixed capacity among the increasing demands (number of comparisons to be made). Each additional comparison reduces the amount of resources available for each comparison and, hence, requires a longer time for all comparisons to be completed. The problem with this model is that the limited capacity can only be used for the comparison process. However, introduction of a concurrent memory load task has been shown to have virtually no effect on the RT (Darley, Klatzky, and Atkinson, 1972).

Another possible model suggests a search through the positive set in which the test item is compared serially to each of the memorized items, and each comparison results in either a match or mismatch. Linear RT functions, as found in the item recognition task, do suggest that subjects use a serial search process whose mean duration increases by one unit for each additional comparison. There are two types of serial search to consider. In self terminating serial search, the test stimulus is compared successively to one item in memory after another, either until a match occurs (leading to a positive response), or until all comparisons have been completed without a match (leading to a negative response). In exhaustive serial search, the test stimulus is compared successively to all the memorized items before a response is made. A self terminating search might require a separate test, after each comparison, to ascertain whether a match had occurred, rather than only one such test after the entire memory set has been compared to the probe. On the other hand, an exhaustive search must involve more comparisons, on the average, than a search that terminates when a match occurs.

The theoretical prediction of RT functions differs for the two models. In an exhaustive search the test stimulus is compared to all items in memory regardless of whether a positive or negative response is required. Therefore, given the equal probability of a negative or positive response, the rate at which RT increases with memory set size is the same for positive and negative responses. This is not the predicted function for the self terminating model. Here, search stops in the middle of the list, on the average, before positive responses, but continues through the entire list before negatives. The result is that as memory set size is increased, the latency of positive responses should increase at half the rate (slope) of the increase for negatives (Figure 8).

A second difference between the two types of search is in the serial position functions for positive responses. Assuming subjects make comparisons in the memorized order, varying the position of the matching item in the list should yield a reaction time function with zero slope for exhaustive models. That is, since every item in the list is compared before the response is made, the response would be made just as quick if the match

116
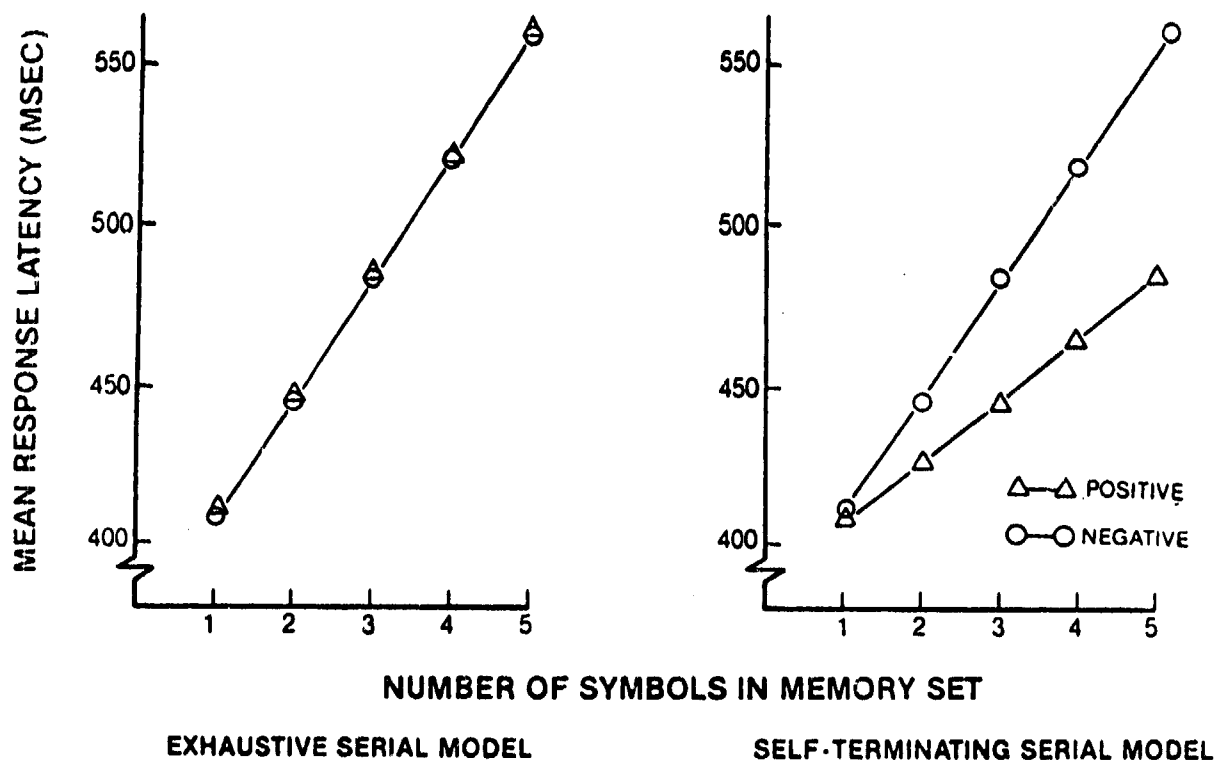
Figure 8. Predicted Reaction Time Functions for Exhaustive
Serial Model and Self-Terminating Serial Model

117

occurred at the end of the list as it would if the match occurred at the beginning of the list. For self-terminating models a match at the beginning of the comparisons process would yield a quicker response than a match at the end of the list resulting in a function with a positive slope.

The serial position curves actually observed in this item recognition experiment were relatively flat (zero slope). This, together with the linearity of the latency functions and the equality of their slopes for positive and negative responses, support the existence of exhaustive search. This does appear to be contrary to common sense and is contrary to subjects reports.

Other Components of RT

The reaction time was defined earlier as the time measured from the onset of the test stimulus to the response. This time is made up of several components which can be related mathematically by the equation:

$$RT = b + a.s \qquad (1)$$

where RT is the mean reaction time, b is the y intercept, a is the slope, and s is the size of the positive set. The slope component of the equation has already been identified as representing the "processing time" (search and decision) unique to that number of items in memory. It is an estimate of the time per comparison and has a value of approximately 38 msec indicating an average scanning rate between 25 and 30 digits per second. Variables affecting the slope of the function have already been described. The other component of the equation is the intercept value of the reaction time versus memory set function. The height of the zero intercept indicates that a large fraction of the RT reflects the duration of processes other than scanning. These processes are believed to represent the basic input/output time. By manipulating different experimental factors, Stérnberg (1969a) identified these processes and arranged them into stages whose durations contribute to the zero intercepts but do not affect the slopes of the functions (Figure 9).

118

Figure 9. Four Processing Stages in Item Recognition [Above
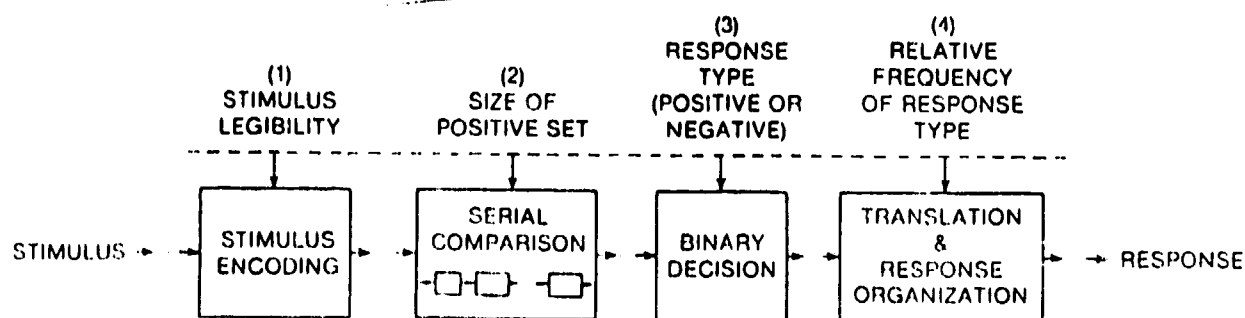the Broken Line are Four Experimental Factors
Believed to Influence the Stages. Vertical Arrows
Show Each Factor Influencing Only One Stage]

119

The first stage involves stimulus encoding and deals with input time. The duration of this stage is affected by the legibility of the stimulus. If the stimulus is degraded or rotated, the additional time needed to encode the stimulus will be reflected in the intercept value of the reaction time. This representation is then used in the serial comparison stage, whose duration increases linearly with positive set size; this is reflected in the slope as discussed previously. In the third stage, a binary decision is made that depends on whether a match has occurred during the serial comparison stage preceding it, the mean duration of the third stage is greater for negative than for positive decisions. The selection and output of a response, based on the decision, is accomplished in a fourth stage, whose duration is influenced by the relative frequency with which a response of that type is required. These last two stage durations, as the first, are also reflected solely in the intercept value. Other factors, of course, may also influence these same stages.

The Sternberg Paradigm in Other Research

Since the task's development and formalization (Sternberg, 1966, 1967, 1969a), it has been subjected to numerous investigation and replication, which has yielded many conflicting results and controversies. Despite the voluminous literature, there have been few attempts to systematically review the great amount of research in this area. One review has been conducted by Sternberg himself (1975), in a well organized albeit subjective article. The other known review was conducted by Hann (1973).

Hann organized the literature according to the type of situational (independent) variable manipulated by the investigators. Thirty distinct independent variables have been identified in the literature and have been collected into seven groups. Varying the memory set size is a feature of all but a few studies since this is one of the basic characteristics of the Sternberg paradigm. RT is the dependent variable for all experiements. The seven categories of variables, as well as some respective studies, are briefly stated.

1. Stimulus category and quality as a variable.

The greatest number of studies reported have been of this type. "Stimulus category" is used in the sense of a formal or conceptual relationship (e.g., digits versus letters, word versus synonym, four sided versus six sided figure, etc.). A typical finding of this group was the more rapid scanning of formally (i.e., physically) similar stimuli, compared to stimuli with associational similarity (Lively and Sanford, 1972; Klatzky, Juolea, and Atkinson, 1971; Naus, Glucksberg, and Ornstein, 1972), also true for same versus different modality manipulations.

2. Stimulus probability and frequency.

In these studies, the probability or frequency of a test item belonging to the positive set was varied (Briggs and Swanson, 1969; Theios et al., 1973). The general conclusion to be reached from these studies is that probability of occurrence of a particular stimulus has an inverse relation to RT in a memory scan task. Whether an item is repeated, specifically cued, or just occurs more often over a series of trials, the results were always a reduction in RT for that item.

3. Temporal variables.

These investigations have manipulated time factors during various phases of the memory search task to study their effect on RT. Varying presentation rate of the memory items seemed to have little or no effect on RT (Burrows and Okada, 1971). However, altering the delay between memory set and test set presentation appeared to effect the memory set encoding process; it was hypothesized that at the shorter delay, comparison is held up until encoding is complete (Connor, 1972).

4. Spatial and numerical separation.

The majority of work in this category has been done by DeRosa, Morin and Associates (Morin, DeRosa, and Stultz, 1967; DeRosa and Morin, 1970; Morin, DeRosa, and Ulm, 1967). It was found from these experiments that when

121

stimuli are organized in some way, such as the well learned properties of a
numerical sequence, the RT is facilitated. On negative trials, the farther
a probe was numerically from the positive set, the faster the RT.

5. Instructional variables.

Several researchers have manipulated independent variables which require
active, intentional processing under the control of the subject, as
instructed by the experimenter. In some experiments, the subject's task
was to mentally remove N items from the positive set (P) so that the number
of items which required a positive response was P-N (DeRosa, 1969; (DeRosa
and Sabol, 1973). Delaying the test probe after presentation of the
deleted items resulted in decreasing RT with increasing delay. Speed ver-
sus accuracy instructions both evidenced strong practice effects (Lively,
1972); however, these effects were noted on the intercept of the RT func-
tion but not on the slope.

6. Test set size.

Manipulations of the test set size has provided additional information
regarding the scanning processes by permitting the decomposition of the
comparison stage into: (1) a retrieval from long term memory followed by
(2) the actual item by item comparison. When there were items common to
both the memory and test sets, the RT dropped as a function of the number
of common items (Briggs and Blaha, 1969; Briggs and Swanson, 1969; Briggs
and Johnsen, 1973).

7. Miscellaneous variables.

Presentation of picture versus letter stimuli to both halves of the visual
field resulted in hemispheric differences in RT (Klatzky and Atkinson,
1971). Picture RTs were faster when processed by the left hemisphere, vice
versa for letter sets. When stimuli were presented to the "slow" half of
the brain for that type of stimulus, the intercept increased but the
comparison rate was unchanged. This additional time was thought to be the

122

interhemisphere transfer time required to get the information to the optimal hemisphere.

Generalizations and Extensions of the Paradigm and Phenomena

Reaction time functions that are approximately linear and increase as a function of set size for both positive and negative responses have been observed in various laboratories with a variety of stimulus ensembles, both auditory and visual. The stimuli that have been used include visual and auditory digits and letters, two and three digit numerals, shapes, pictures of faces, drawings of common objects, words of various lengths, colors, and phonemes (e.g., Burrows and Okada, 1973; Chase and Calfe, 1969; Clifton and Tash, 1973; Foss and Dowell, 1971; Hoving, Morin, and Konick, 1970; Swanson, Johnsen, and Briggs, 1972). The slopes of the different ensembles are not the same but differ systematically in an orderly way. The RT functions have been observed to remain linear and parallel in studies with positive sets containing up to 10 letters, (Wingfield and Branca, 1970) and up to 12 common words (Naus, 1974).

The phenomena have been observed in people of various ages, ranging from children to elderly adults, and in normals, alcoholics, schizophrenics, and brain damaged mental retardates. For some of these groups, the slopes and/or intercepts of the RT functions are elevated relative to those of young adults; for example, aging and mental retardation both appear to produce increased slopes (Anders, Fozard, and Lillyquist, 1972; Harris and Fleer, 1974). Children as young as eight produce RT functions with higher intercepts, but the same slope as young adults (Hoving et al., 1970; Harris and Fleer, 1974). Also, except for differences in the value of the y-intercept, schizophrenics and alcoholics look surprisingly similar to each other and to normals.

Finally the effect of extended practice in the item recognition task should be considered. The effect seems to depend on details of the procedure. Several studies have shown that when subjects practice with the same fixed sets over many days, the RT functions become flatter and negatively accel-erated. This is particularly true if members of the ensemble are

123

consistently associated with particular responses; so that a stimulus that is in any positive set for a subject can never be in any negative set, and vice versa (Ross, 1970; Kristofferson, 1972b). On the other hand, when sets are changed either from trial to trial or from session to session (Kristofferson, 1972a), and stimuli are not consistently assigned to particular responses, extended practice seems to have virtually no effect on the phenomenon. The effect of practice also seems only to affect the zero intercept, not the slope (Kristofferson, 1972a).

RELIABILITY

The item recognition task has been tested for stability of scores for its possible inclusion in a battery of Performance Evaluation Tests for Environmental Research (Peter), (Carter et al., 1980; Carter and Krause, 1983). If a test is to be used for drug or environmental research, it must be administered repeatedly to the same subjects in a baseline condition and in the novel environment. It would be desirable for a test to provide unchanging scores in the baseline because any change associated with repeated measurement would be confounded with changes of performance due to the environment.

In the Carter et al. (1980) study 21 male subjects performed the item recognition task with positive set sizes of one to four digits which were presented for 1 second per item. Each session included 10 trials for each memory set size with half of these trials requiring a positive response and the other half a negative response. Digits were chosen at random, and were different on each day, but were the same for all subjects on any particular day. Testing was conducted once each day for 15 consecutive weekdays. The test sessions lasted about 15 minutes per subject per day. Data was obtained for mean RTs for positive set sizes, slope of mean RT versus set size, intercept of mean RT versus set size, and percent error.

The intercept score did not change appreciably during the experiment, slopes decreased with practice until the third day and response times stabilized after the fourth session.

The intersession reliabilities of slopes and intercepts indicated the
degree to which the scores represent enduring abilities (remain in the same
relationship from day to day). The intersession reliabilities for both
slope and intercept scores were found to be uniformly low. According to
Carter et al., (1980), the poor reliabilities cast doubt upon the potential
of these scores for measurement of individual differences and they would
make the test relatively insensitive to environmental effects. However, it
should be taken into consideration that very few trials per memory set size
were given during each day in this study (five positive and five negative
trials). It is not surprising to find low reliability scores for the slope
with so few trials. In contrast, the reliabilities of the RTs from which
the slopes are calculated were relatively high, being generally greater
than .70. Thus, RT was stable for each of the four memory set sizes, from
the standpoint of reliability, after the fourth session.

VALIDITY

The item recognition paradigm developed by Sternberg (1966) is a memory
search task which utilizes error free reaction times to determine processes
of retrieval and comparison in short term working memory. The slope of
these reaction time functions is taken as a measure of the rate of search
through short term memory, and the intercept is interpreted as the time
required for stimulus processing and response formulation (Sternberg, 1966,
1975).

Results obtained with the item recognition paradigm have been duplicated in
a number of experiments demonstrating that the phenomenon is relatively
robust, and that the estimated scanning rate is remarkably invariant across
subject populations and practice. The most general observation is that
investigators have found memory scan to be a serial process. That is,
regardless of other variables, RT was always an increasing function of
positive set size. Also, with a few exceptions (e.g., Klatzky et al.,
1971; Holmgren, 1970), violations of the assumption of nonoverlapping
stages and the assumption of pure insertion have not been found necessary
to explain the data.

Effects of duplication of items in the list, their serial positions, and
the relative frequency with which they are tested, have led investigators
to support different models of memory scanning. Roughly, twice as many
investigators have supported the exhaustive scan theory than have favored
the self terminating search interpretation; however, the latter group is
sizeable. Also, another group of researchers, as large as the self ter-
minating group, has found neither explanation to be wholly satisfactory,
favoring instead various combinations of the two theories.

In summary, this memory search task does appear to be diagnostic of the
processes involved in retrieval and comparison of items in short term work-
ing memory as evidenced by the slope of the RT function. To a lesser
extent, this task is also diagnostic of the time required for stimulus
encoding and response formulation as evidenced in intercept scores. The
underlying models of search processes have not yet been clearly estab-
lished; however, given the purpose of the UTC-PAB, the underlying model
describing memory search is not of critical importance.

SENSITIVITY

Various modifications of the Sternberg memory search task have been used
frequently in environmental research. The intent of this research is not
always the same. This section has been divided into two classes of envi-
ronmental research in which the Sternberg task is used as a measure of
short term memory performance. These classes are: (1) drugs and (2) work-
load, which is further broken down into physiological and dual task
research. Representative studies from each class and their findings will
be described to determine the sensitivity of the task to manipulations of
these environmental factors.

Drugs

By examining the slopes and intercepts of reaction time versus set size
functions, in drug treatment and placebo conditions, the locus as well as
the presence of drug effects can be determined. In one study, the memory
search task was used to evaluate the dose response relationship between

126

elemental mercury exposure and short term memory functioning (Smith and Langolf, 1981). Set sizes of two, three, and five digits were presented using the fixed set visual procedure to 26 male workers in two mercury cell chloralkali plants. Workers were tested twice at a three month interval. Intercept, memory scanning time, and effect of response type were measured as dependent variables. Intercept was not significantly related to any of the four mercury exposure indices. However, memory scanning time was significantly related to all four indices and the effect of response type was significantly related to the two lower doses. The authors concluded that chronic exposure to mercury may have a detrimental effect on memory scanning time and that the locus of this effect exists in the central nervous system.

In another application, Osborne and Rogers (1983) attempted to determine the effect of various combinations of alcohol and caffeine on human reaction time. In this application, the Sternberg paradigm was used to help determine which processing stages are most effected by the drugs. Set sizes of one to four letters were visually presented to eight subjects in random order via the fixed set procedure. The results showed no significant differences in the slopes of the various alcohol/caffeine combinations; however, significant differences were obtained with the intercept values. These results led the authors to conclude that these drugs affect the peripheral stages in the Sternberg information processing model.

Two antidepressant drugs, amoxapine and amitriptyline, were given to depressed outpatients whose reaction times on the memory search test were measured before and after treatment (McNair, Kahn, Frankenthaler, and Faldetta, 1984). Using a positive set size of from one to six digits, specific digits, series lengths, test digits, and position of positive test digit in the preceding series were randomly generated. A significant increase in speed of performance was associated with amitriptyline, about 7 percent faster. Amoxapine neither impaired nor facilitated performance on the task.

Roth, Tinklenberg, and Kopell (1977) used the Sternberg tasks to elicit event related potentials (ERP) which were used to compare the effects of

ethanol and marihuana. Twelve subjects were tested on three separate days
1 hour after ingestion of one of the drugs. On each trial, one to four
target digits were presented consecutively followed by a probe digit. Each
target set size, each portion of the probe in this target set sequence, and
in set and out of set probes were randomized. ERP measures were then
taken. P300 amplitude showed both a drug effect and a set size effect.
Both drugs differed significantly from the placebo but not from each
other. Marihuana increased overall RT for each set size by about 75 msec.

The Sternberg memory scanning task was one of three tasks given to 18 sub-
jects after receiving 10 mg of metamphetamine, 100 mg secobarbital, and a
placebo on separate days (Mohs, Tinklenberg, Roth, and Kopell, 1980).
Tests were given before treatment and 50 minutes following drug administra-
tion. Subjects were given a series of trials lasting a total of 20 min-
utes. At the start of each trial, a new memory set of one to four digits
was visually presented (V-VS). Neither drug significantly affected per-
formance on this task. RT did increase linearly with set size and there
were fewer errors (12 percent). Thus, metamphetamine and secobarbital do
not affect short term memory.

The results of the described studies provide evidence that tasks, for which
well developed cognitive theories exist such as the Sternberg memory search
task, make it possible to study the performance of specific stages or com-
ponents of performance. Because of this property, they are well suited to
application in the field of behavioral toxicology.

Dual Task

The Sternberg task is also particularly appropriate for the purpose of
localizing dual task effects within stage theory. It is thought that the
Sternberg task may be sensitive to the memory load the individual is under
while performing a separate, primary task. The positive set would be a
sample of the individual's total memory load which would then be eval-
uated. When the Sternberg task is used as a secondary task, it would be
hypothesized that the slope of the function would be a measure of primary

128

task memory load and the intercept would be an estimate of the secondary task interference with the primary or vice versa.

Reaction times in the Sternberg task were used to localize the divided attention effect (less proficient performance under dual than under single task conditions) within the stage model (Briggs, Peters, and Fisher, 1972). A tracking task was used as the primary task as it was expected to load across all stages of information processing equally. The Sternberg fixed set (one, two, or four items) procedure was auditorily presented to the subjects as the secondary task. The results showed a dual task effect of intercept only. Briggs et al. (1972) concluded that when loading is broadly based across stages, then the primary divided attention effect seems to be manifested rather early in the processing of information by the human, such as in the encoding (input) stages.

Spicuzza, Pinkus, and O'Donnell (1974) have also used the memory search task as a secondary task to measure the effects of Manual Flying Work-load. Both auditory and visual presentations of the fixed set procedure were used with memory sets of one, two, three, and four letters. The subjects were given one of two simulated flying missions as the primary task. The authors concluded from their results that standard Sternberg methods of scoring appear to yield consistent and interpretable data with predominantly linear trends.

Crosby and Parkinson (1979) investigated pilots' skill levels by measuring performance of instructor pilots and student pilots in a dual task paradigm, combining a ground controlled approach (GCA) as the primary task and memory search as the subsidiary task. Between groups differences on the search task were restricted to the intercept of the function. It was concluded that the effect of experience on the type of flight task examined was to reduce the processing demands of encoding or responding. Also, dual task performance discriminated between student groups, differing in only four weeks of training, suggests that the dual task paradigm has considerable potential value in providing an objective measure of flight proficiency.

Wetherell (1981) used the memory search paradigm as one of a battery of secondary tasks to measure the mental load imposed by driving under standard conditions. Subjects heard series of four or eight random digits from the range 0 to 9 at the rate of one digit per second. The only significant finding with this task was that the proportion of sequences correctly recalled by males decreased significantly, while the effect was similar but not significant for females.

Event Related Potentials (ERPs)

A final use of the Sternberg memory search task is to examine psychophysiological responses (i.e., P300 latency). This task is ideal because it is a more complex task in which the stimulus events are readily discernable and performance measures are maintained at acceptable levels. By recording brain potentials to positive and negative test stimuli while varying the number of items, it may be possible to observe differences in waveform as a function of stimulus class or complexity. In an early experiment, a significant enhancement of the P300 (late, positive) component was observed for positive letter presentation in item recognition tasks. The difference between negative and positive probes increased with positive set size, and RTs were significantly longer for negative stimuli (Gomer, Spicuzza, and O'Donnell, 1976).

Late positive components have been used with the memory search paradigm to try to define the underlying models of the search task. Brookhuis, Mulder, Mulder, Gloerich, VanDellen, VanDerMeere, and Ellerman (1981) measured amplitude and latency of late positive components together with RT on the memory search task. The visual varied set procedure was used with a memory load size of one to four characters. The RT data indicated a self terminating search process while the P300 data suggests an exhaustive search process. Several possible solutions for the results are suggested.

Adam and Collins (1978) used digits of set sizes 1, 3, 5, 7, 9, and 11 and recorded brain potentials. Results supported a serial and exhaustive search. P300 latency increased with set size up to size seven with an average search time of 22 msec per set item. With set sizes 9 and 11, the

results indicated large individual differences and also a break in the correlation between RT and ERP latencies.

The effects of age differences on memory search have also been measured by ERPs. In one study, the amplitude and latency were not significantly different for young and elderly subjects, but the RT was significantly slower for older than younger subjects (Ford et al., 1979). In another study, however, the amplitude of the P300 increased significantly with set size, and younger subjects had significantly larger P300 amplitudes than older subjects. These effects matched the RT functions (Pfefferbaum et al., 1980).

As evidenced by the discrepancies of the results for the studies described, the validity of the event related potential is questionable until further definitive research is performed.

TECHNICAL DESCRIPTION

The six versions of the UTC-PAB memory search task will share a number of common specifications. In all versions, the positive set items will be randomly selected from the 26 English alphabet characters. However, no items which are acoustically confusing will be used in the same positive set. The negative probe letters used with a specific positive set will be randomly selected from the remaining alphabetic characters with the restriction that none will be acoustically confusable with any member of the positive set. In all cases, trials will consist of 50 negative probes and 50 positive probes presented in a random order.

The visual versions of the task (V-FS, V-MS, V-VS) will use upper case alphabetic characters. Subjects will view the CRT from a distance of 60 cm. Positive sets will be presented simultaneously on a line approximately one-third of the distance from the top of the screen. Probe letters will be centered on a line one-half the distance from the top of the CRT. Letter size for all stimuli will be .5 cm wide by .7 cm high.

The fixed set versions (V-FS, A-FS) will begin by presenting the positive set for subject inspection on the CRT. In A-FS, the set will also be spoken at a rate of one character per second during inspection. When the subject has memorized the list, the subject will press either of the two response keys which will remove the positive set from the screen, terminating the inspection period and initiating the trial. One second after the subject terminates inspection, the first probe letter will appear. Succeeding probes will be presented 300 msec following the response to the previous probe. Probe letters on the V-FS procedure will remain on the screen until the subject responds. In either the V-FS or A-FS version, if the subject fails to respond to the probe within 3 seconds, a 1000 Hz tone will sound for 300 msec, the next probe will be presented, and the presentation will be scored as a "response failure." No reaction times will be recorded in these cases. The mixed set versions of the task (V-MS, A-MS) will have timing and response deadline characteristics identical to the fixed set versions. The varied set versions will also be identical to the fixed set versions with the exception that the time available for observing and encoding the positive sets will be fixed at 1 second. Once this period has elapsed, the probe stimulus will be automatically presented.

Trial Specifications

The chronological series of events for the fixed, mixed, and varied versions for each trial are established as follows:

(1) Fixed Set Versions: (a) positive set inspection time, terminated by onset of subject's start response, (b) first probe onset, 1 second following subject's start response, (c) reaction time onset of probe to onset of subject's choice response, and (d) response probe interval fixed at 300 msec (onset of choice response to onset of probe). A trial consists of the presentation of one positive set followed by 100 probes.

(2) Mixed Set Versions: (a) positive set inspection time, terminated by onset of subject's start response, (b) first probe onset, 1 second following subject's start response, (c) reaction time onset of probe to onset of choice response, and (d) response probe interval fixed at 300 msec. Ten

132

probes follow each positive set. A new positive set appears 300 msec
following 10th choice response. A trial consists of 10 positive sets
followed by 10 probes.

(3) Varied Set Versions: (a) positive set inspection time fixed at
1 second, (b) probe stimulus onset 300 msec following offset of positive
set, (c) reaction time onset of probe to onset of choice response, and (d)
new positive set appears 300 msec following onset of previous choice
response. A single probe follows each study set. A trial consists of 100
study sets followed by one probe.

DATA SPECIFICATIONS

A separate data record will be stored for each trial. Each record will
contain the specific positive sets and all probes used in a trial. From
the start of every trial, certain times in msec shall be recorded. These
are: (1) trial start, (2) onset of study set, (3) offset of study set, (4)
onset of probe item, (5) onset of subject response to probe, and (6) onset
of deadline alarms.

From these time measurements and data, statistics can be calculated and
various RTs can be computed. These, in turn, can be used to determine
slope and intercept values of the RT versus positive set size functions.
The summary statistics suggested include: (1) mean positive set inspection
time for both fixed and mixed versions, (2) mean correct RT and standard
deviation to probe items, (3) mean correct RT and standard deviation to
positive probe items only and to negative probe items only, (4) total trial
duration, (5) number and percent of response failure errors, (6) number and
percent of incorrect response errors, and (7) number and percent of total
errors.

TRAINING REQUIREMENTS

The instructions should be read to the subjects before the start of the
training trials. In all versions, subjects are instructed to respond to
the probe stimuli as quickly and accurately as possible. However, accuracy

133

is emphasized and subjects should attempt to keep error rates below 5 percent in any trial. In the fixed and mixed set versions where the inspection period for the positive set(s) is determined by the subject, subjects should be told to take only enough time to insure representation of the positive set in memory. Precise training times for the six versions of this task have not been determined. However, generalizing from other similar research, major practice effects are eliminated with four training sessions composed of 7 to 16 trials with each positive set size.

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

The memory search task consists of two parts. In the first part of the task, you will be memorizing a small set of letters from the alphabet. This is called the "memory set." In the second part of the task, you will see a series of letters presented one at a time. Your task is to decide whether each letter is one of the letters in the memory set. If a letter is one of the memory set items, you press the "yes" key; if it is not one of the memory set items, you press the "no" key. The object of the task is to respond to the letters as quickly as possible without making any errors. Respond as fast as you can to the letters, but if you find

yourself making errors, slow down. You should try to respond correctly to every item.

There will be either one, two, four, or six letters in the memory set. On some trials, you will have as much time as you need to memorize the letters in the memory set. On other trials, this time will be set for you. It should take you not more than 15 to 20 seconds to commit the items to memory. The actual letters in the memory set will be different on each trial, so you'll have to memorize a new set at the beginning of each trial. On certain trials only one probe letter will follow the memory set, on other trials 10 probes or 100 probes will follow the memory set. Also on some trials the probe letters will be presented acoustically, while on other trials they will be presented visually.

SPATIAL PROCESSING TASK (UTC-PAB TEST NO. 10)
(SPATIAL ORIENTATION/ROTATION SHORT TERM MEMORY)

PURPOSE

This task is designed to examine the subject's ability to mentally rotate a series of histograms prior to making a same/different judgement about them. The task taps visual short term memory, since the standard and test stimuli are presented successively rather than simultaneously.

DESCRIPTION

The subject will be presented a series of histograms one at a time. He must determine whether the second histogram of each pair is identical to the first. He will indicate his answer by either pressing a button labeled "same" or a button marked "different" on a two key response box. Task loadings are varied by presenting a two bar standard stimulus with the test stimulus in the zero degree orientation for low loading; a four bar standard with the test stimulus in the 90 or 270 degree orientation provides moderate task loading; and a six bar standard with the test stimulus in the 180 degree orientation provides high task loading.

BACKGROUND

This version of the spatial processing task is from the criterion task set (CTS) (Shingledecker, 1984). The CTS version is in turn derived from an earlier task used by Chiles, Alluisi, and Adams (1968). In the original task, the subjects were shown a standard stimulus and then a pair of test stimuli. The subject's task was to decide if one, neither, or both of the test stimuli were identical to the standard stimulus. The standard was presented for 5 seconds and each test stimulus was presented for 2 seconds. One second elapsed between each successive presentation. The quality of the test stimuli was degraded by the introduction of "noise" in the pattern. Noise was defined as a random state change of a matrix cell (i.e., making it white when it was originally black or vice versa).

The CTS version of the task is somewhat different. A standard stimulus oriented at zero degrees is presented. After a pause, a single test stimulus is presented in an orientation of 0, 90, 180, or 270 degrees. The figure may be the same as, or different from, the standard stimulus. The prime similarity between the two versions (CTS and Chiles et al., 1968) of the task is the type of stimuli.

The current experimental task is taken from the CTS battery (Shingledecker, 1984). Individual tasks in the battery were designed to place specific and selective demands on the capabilities of the human subject. The capabilities (or resources) chosen were hypothesized to be prime components of a variety of more complex human behaviors typically occurring in both military and civilian workplaces. During the development phase of the spatial processing task, all elements of the test (e.g., number of bars and test stimulus orientation) were combined factorially. Levels in the current task represent three levels from the development phase which were shown to have reliable and statistically significant differences between them. Although in a strict experimental design sense there is a confounding of orientation with number of bars in the stimulus (since not all orientations occur with each number of bars), the purpose of the task is to produce reliably different loading levels. The different loading levels are, therefore, the important aspect of the task rather than the interrelationship of the task's factors. The purpose of the task must, above all, be sensitive to the different loading conditions.

The structure of the model posits three stages of processing and associated resources: perceptual input, central processing, and response output. The tasks were selected from the literature of cognitive and psychomotor performance which coincided with the various combinations of input, processing, and output modes in the model. These tasks were then, in turn, validated and different levels of task loading were determined. Thus, the spatial processing task used in the UTC-PAB was designed to load spatial memory and matching abilities in the model.

In the Chiles et al. (1968) task, the stimuli were all six bar histograms, with each bar ranging in height from one to six units. No two bars in the

137

same figure could be identical in height. The Shingledecker (1984) CTS stimuli have either two, four, or six bars.

The differences between the two tasks are great enough to make generalization from one to the other questionable. In the Chiles et al. (1968) task, the primary loading is a memorial one. The standard must be maintained in a memory store for comparison purposes; since there are two separate test stimuli, the test stimuli must also be stored. A minimum of two separate comparisons must be made, with the intermediate results of each comparison maintained in memory as well. The figures are not manipulated by the subject in this task, only compared.

In the CTS version the standard must be maintained in memory, but the test stimulus does not. In all but the two bar histograms, the test stimulus must be mentally rotated prior to the same/different judgement (see Cooper and Shepard, 1978 regarding mental rotation and same/different judgements). Thus, the primary loading for the moderate and high difficulty levels of the task (the low level task is excluded here since the test stimulus is always in the same orientation as the standard) would appear to be a spatial transformational one.

The Chiles, Alluisi, and Adams (1968) task on which this test is based is somewhat different, both in structure and intent. In that task, the subjects were shown a target pattern, whose basic construction was identical to the six bar histograms in the CTS task. They were then shown two test stimuli in succession. However, prior to display of the test stimuli, some level of noise was introduced by changing the state of certain cells in the matrix (i.e., turning them on when they should be off, or vice versa). The subject's task was to indicate whether the first, second, or neither test stimulus was identical to the standard stimulus. The CTS version does not introduce noise into the matrix, nor does it ask the subject to make judgements about a pair of test stimuli.

The original version of this task was created by Fitts et al. (1952) and the general paradigm is referred to as the Fitts Histogram procedure. In this eariler work, Fitts and his collegues presented a single histogram to

138

their subjects as a standard, followed by six rows of eight simultaneously presented test stimuli. The subject's task was to select the test stimulus from each row that was identical to the standard. Some of the stimuli were created in the same fashion as those in the current study, using six bars with lengths from one to six units. Others were created as the figure and its mirror image, joined at the midline. And finally, a third group was composed of two iso-oriented repetitions of the pattern. In general, Fitts found that response time was fastest for random stimuli, and slowest for constrained stimuli (i.e., stimuli with bars chosen without replacement from the population of possible heights). In addition, symmetrical stimuli were identified most quickly.

The type of task used in the current experiment probably falls into the category of spatial transformation as defined in Lohman's 1979 survey and reanalysis of the correlational literature on spatial perception. More specifically, the task probably requires visualization (Vz) ability. Vz tasks involve the mental reorientation (e.g., mental rotation) of complex figures or designs prior to making judgements about those figures. The complex figures in Vz tasks are most often two dimensional representations of three dimensional objects. Sometimes the figures are plane polygons as in the current study. Because the tasks involve the manipulation of a great many figural points and planes, Vz operations are often characterized by relatively slow performance. This type of slow performance is typical of Kosslyn and Shwartz's (1977) CRT model of mental imagery, where mental rotations and manipulations are the result of point by point transformations of the mental image by the subject. A simpler (and somewhat faster) type of spatial transformation is labeled spatial orientation (SO). Rather than mental rotation of the stimulus figure, the subject typically imagines observing the figure from a new vantage point or perspective. It is unlikely that SO operations would be used for the current task, since the histograms are purely and obviously plane figures, rather than two dimensional representations of three dimensional objects (as the figures were in Shepard and Metzler's 1971 study where Vz strategies were most often used). The final level of Lohman's hierarchy of spatial factors and processes contains factors which may apply to the current task. Since the task must be performed under time constraints, the spatial orientation test

139

will probably be affected by the perceptual speed (Ps) dimension which is best described as the speed of matching stimuli, and closure speed (Cs) which is the speed of matching incomplete or distorted stimuli with representations already in memory. Lohman's hierachy is presented in Figure 10.

The stimuli used in this study were originally developed by Fitts and his colleagues and were called constrained figures. This meant that each bar in the histogram was selected from a population of all possible bar heights without replacement. Therefore, no two bars in the figure can have the same height. Fitts also used random figures. The bar heights for these figures were chosen at random, so it was possible for two or more bars in a figure to have the same height. Generally, Fitts and his coworkers found that detection times for the random figures were faster than for the constrained figures.

RELIABILITY

Kennedy and his colleagues (1985) used the Fitts Histograms as a marker test during the development of a microcomputer based repeated measures test battery. They found a test-retest reliability for the task of 0.90. Using the Spearman Prophecy formula, they estimated the reliability of a 3-minute version of the test to be 0.93. The test, in the Kennedy study, was administered as a paper and pencil test. This type of test tended to stabilize more slowly than the same test in computer based form. Therefore, any generalizations must be made with caution. The Chiles et al. (1968) task has a split half reliability of 0.75.

VALIDITY

The Fitts Histogram test correlated 0.71 with the Klein and Armitage task (a simultaneous dot pattern comparison test included in the UTC-PAB) in the Kennedy et al. (1985) study. Previous research has shown that the Klein and Armitage pattern comparison test loads on spatial factors. Kennedy and his coworkers performed a factor analysis on the tests in their battery (again, these results should be interpreted with caution since there were only 20 subjects and 11 tests) and isolated four factors. The Fitts

140
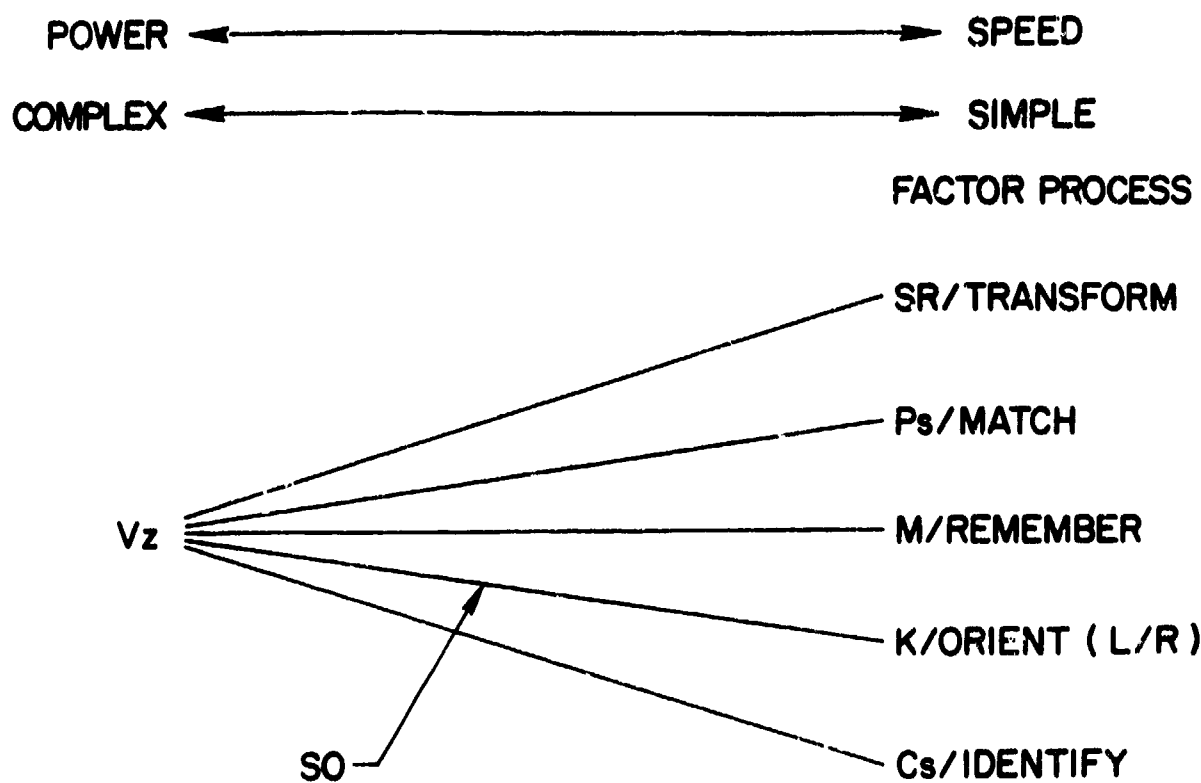
Figure 10. Representation of the Relationships Between the Various
Spatial Factors and Abilities (After Lohman)

141

Histograms loaded on the same factor as the Manikin test (a test loading on Lohman's SO factor), code substitution (loading on SR), and the Klein and Armitage task (also Lohman's SR factor). Fitts Histograms also loaded on a factor which appeared to be a motor control factor (this can probably be attributed to the fact that the test was administered as a paper and pencil test). One rather interesting fact: one factor was representative only of the computer based tasks and not their paper and pencil counterparts. This suggests that there might be fundamental differences in the strategies or behaviors used by subjects in addressing different versions of the same test.

SENSITIVITY

Sensitivity to Intrusive Agents and Factors

No research has been completed with Fitts Histograms examining the effects of drugs, toxic agents, or environmental stressors. Similar research, however, has been perfromed on tests which load on the same spatial factors as the Fitts task. The Manikin test has been shown to be sensitive to the effects associated with diving to extreme depth (e.g., 600 meters) (Lewis and Baddeley, 1981; Logie and Baddeley, 1983). The Klein and Armitage test has been demonstrated to be sensitive to cyclical variations in cerebral hemisphere arousal (Klein and Armitage, 1979). Chiles, Bruni, and Lewis (1969) and Chiles, Alluisi, and Adams (1968) used a task like the Fitts Histograms in studies of long term vigilance and social interaction during isolation.

TECHNICAL DESCRIPTION

The histograms will be composed of bars one to six units in height. In any single histogram, no two bars will be identical. The bars will be separated from adjacent bars by a gap equivalent to a single bar's width. Each histogram will be presented with a horizontal line at its base and a number to designate its presentation position (i.e., a one if the histogram is the standard stimulus, or a two if the stimulus is the test figure). All standard stimuli will be presented in the zero degree orientation (i.e., with

the histogram bars extending above the horizontal line); the test stimuli
will be presented in the zero degree (for the two bar stimulus), 90 and
270 degree (for the four bar stimulus) or 180 degree (for the six bar
stimulus) orientations.

The task is performed in 3-minute trials. The standard is presented for
3 seconds, followed by a 1 second pause. Presentation duration for the
test stimuli varies with the number of histogram bars: a maximum of 1.5
seconds for the two bar stimuli, 2.5 seconds for the four bar stimuli, and
3.5 seconds in the six bar condition. The subject's response must be made
between test stimulus onset and offset. Responses are made on a two key
response box with one key labeled "same" and one key labeled "different."

Trial Specifications

Each presentation during the 3-minute trial consists of the following
events: (a) the standard stimulus is presented for 3 seconds; (b) the
screen clears for 1 second; (c) the test stimulus is presented for 1.5 to
3.5 seconds (dependent on the number of bars in the histogram); (d) if the
subject makes a response before the end of the test stimulus presentation
period, the screen clears until the end of the period; (e) during the
training trials feedback is presented; (f) the next trial is presented.

DATA SPECIFICATIONS

The program generates and records a reaction time for the response to each
test stimulus. In addition, a response code indicates whether the response
was correct, incorrect, or terminated by the deadline.

A variety of summary statistics are computed including: (a) length of
presentation; (b) number of presentations; (c) number correct; (d) percent
correct; (e) percent presentations terminated by deadline; (f) percent
incorrect; (g) percent total errors (including deadlines and incorrect);
(h) mean correct reaction time; (i) median correct reaction time; and (j)
standard deviation of the reaction time. Hard copy of the data and summary
statistics is also available.

TRAINING REQUIREMENTS

As a first step, subjects should be read the instructions. After the
instructions, the subjects should receive at least a 3-minute trial at each
level of difficulty in order to achieve stable performance. During the
training periods, there is a 15-second response deadline; there is also
feedback to the subject.

It is important that the subjects perform the task in the fashion it is
described in the instructions, (e.g., as quickly and as accurately as pos-
sible). If the experimenter feels that the subject does not understand the
instructions or the task, or is performing incorrectly, additional instruc-
tion and test trials may be administered.

To summarize, the training phase for this test should consist of the fol-
lowing steps:

1. Read the instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure
    that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects
    require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run
    over several sessions on this test, one may omit the practice
    trials after the first session.

INSTRUCTIONS TO SUBJECTS

In the Spatial Processing task, a series of bar graphs, or histograms, are
presented one at a time. Your task is to memorize the shape of the first
of the two histograms, and then decide whether the second histogram is the
same shape or a different shape than the first. The first histogram is

labeled with a "1" and the second with a "2" so that you can keep them straight. Always memorize the shape of the first histogram and make a same/different response when the second histogram is displayed. "Same" and "different" responses are made on the left and right keys of the keypad.

There are three versions of the task. In the first version, the histograms are composed of only two bars and the second histogram in the pair is oriented in an upright position. In the second version of the task, the histograms contain four bars and the second histogram in the pair will appear rotated on its side, either to the left or right. The third version has six bar histograms with the second histogram in an upside-down orientation. The first histogram in each pair will always be presented in an upright position.

You control when the task starts by pressing any of the response keys. Memorize the shape of the first histogram and respond either "same" or "different" to the second. The first histogram will be erased as soon as you respond and the next pair of histograms will start. Try to respond as quickly and accurately as possible. Go as quickly as you can, but if you start making errors because you are rushing your decision, slow down. Data collection lasts for 3 minutes from the start of the trial. After 3 minutes the task will automatically stop and the screen will go blank.

MATRIX ROTATION TASK (UTC-PAB TEST NO. 11)
(SPATIAL ROTATION SHORT TERM MEMORY)

PURPOSE

The purpose of the Matrix Rotation Task is to assess the subject's facility for spatial rotation. Spatial rotation, also known as spatial transformation, is one component of spatial orientation. This task also evaluates short term perceptual memory.

DESCRIPTION

The computer presents a series of 5 by 5 cell matrices, one by one, on the center of the display. Each matrix has five illuminated cells. After a pause, the screen blanks and a second matrix is presented. The subject is required to determine if the second matrix is identical to the first. Responses are made on a two key response box.

A matrix is considered to be identical only if it is a 90 degree rotation of the standard (i.e., first) matrix. Successive test matrices are never presented in the same orientation.

BACKGROUND

The matrix rotation task used in this UTC-PAB test is based on tasks from Phillips (1974) and Damos and Lyall (1984). In the Damos and Lyall study, the stimuli were composed of a 5 by 5 matrix with five illuminated cells. In the Phillips study, matrices were four, six, or eight cells on a side; the matrix grid was not visible. Damos and Lyall did not specify the physical size, makeup, or configuration of their stimuli beyond the dimensions of the parent grid and number of illuminated cells.

Several important differences exist between the Damos and Lyall stimuli and those used in the other spatial tasks in the UTC-PAB, which are worth noting. The first is the number of filled (or illuminated) cells. In the

other studies, the proportion of filled cells was at or above 50 percent of the cells in the matrix (Phillips, 1974; Fitts et al., 1956; Klein and Armitage, 1979; Ichikawa, 1981). In the Damos and Lyall study, only 20 percent (5 of 25) cells are filled. If a large proportion of the cells in the matrix are filled, there is a greater likelihood that patterns of contiguous cells will be formed. In matrices with a lower proportion of cells filled, it is more likely that cells will be isolated within the matrix (e.g., have no filled cells abutting them). This tends to make the pattern more difficult to memorize and manipulate; it is easier to memorize patterns when the components are unambiguously associated or related in some way.

The second issue to consider is related to the first. If filled cells are isolated within the matrix, those cells must be dealt with as individual figures, rather than as part of a larger entity. This makes the figure more complex. The effect of this increased complexity will be dealt with below.

The nature of this task implies that it largely requires spatial abilities. One of the most useful definitions of spatial ability is presented in the work of Lohman (1979). Through an extensive reanalysis of the correlational literature on spatial abilities, Lohman identified three primary factors of spatial skills. The highest level skill was called visualization (Vz). Vz tasks involve the mental reorientation of a complex figure or pattern in mental space. An example would be imagining the letter "R" rotating slowly into an upside down position. A second spatial ability, located lower in the hierarchy, is called spatial orientation (SO). This ability also involves mental rotation, but this time it involves reorientation of the observer's viewpoint rather than the object being viewed. Using the letter "R" again as an example, SO tasks would require the subject to imagine what the letter looks like from the back. The third ability in Lohman's model has been labeled spatial relations (Sr). Spatial relations can best be thought of as the ability to solve spatial problems rapidly, regardless of the means used in solving the problem. See Figure 10 for a representation of Lohman's model.

147

The problem of complexity, mentioned above, comes into play at this point. Tasks which are located high in the hierarchy (such as Vz tasks) are quite difficult. As difficulty increases, speed of task execution decreases. Thus, adding difficulty to the task (e.g., by decreasing the number of filled cells) decreases speed still further. These factors may render comparisons between the UTC-PAB version of the test and other spatial matrix-based test difficult to interpret. That is, the differences between the UTC-PAB implementation and earlier versions of the test may be qualitative rather than quantitative.

Other differences between the task as implemented in the UTC-PAB and its original form may have implications for subjects' performance. In the majority of the parent tasks forming the basis for the UTC-PAB tests, stimuli were typically of dot in matrix construction. The spatial tasks from the present battery, however, are filled cells. The difference in appearance between stimuli with dots in a matrix cell and those with completely filled cells is substantial, even though the same amount of information is conveyed in both stimuli (Royer, 1981). In fact, as Royer found, there may be performance differences between stimuli composed of different design elements.

In his study, Royer used figures composed of two different elements, dots or diagonal line segments (which he termed diagonolinear). Reaction times to the figures composed of diagonals was always slower than to the dot patterns. Royer also generated different elements for pattern development: rectilinear elements (which were orthogonal lines drawn between two filled cells), and block elements (which had each cell completely filled). The differences in appearance between the four types of patterns is striking, although they all contain the same amount of symmetry information (Figure 11).

Another difference between the tests in the UTC-PAB and their source tests is the method of presentation. Some of the parent tests were presented in paper and pencil form. There is some indication (Kennedy et al., 1985) that tests presented in this form show different patterns of performance stability than tests which are computer based.

148

PATTERN ELEMENT TYPE

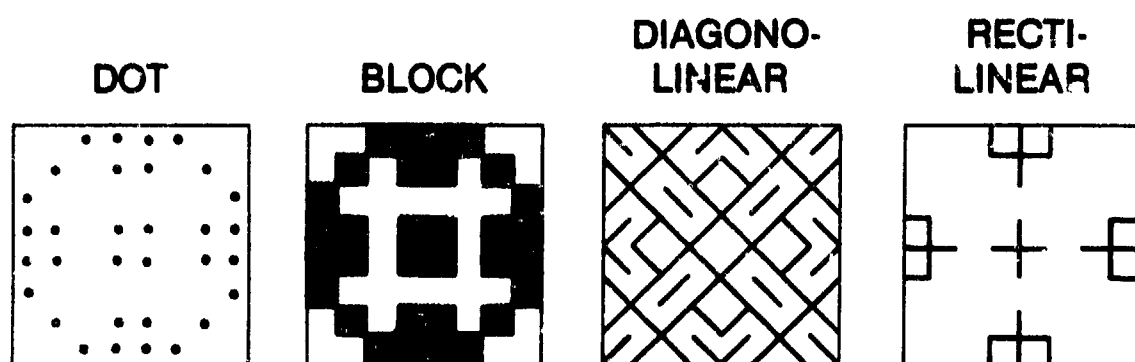DOT   BLOCK   DIAGONO-LINEAR   RECTI-LINEAR

Figure 11.   Examples of the Different Types of Cell Elements and Symmetry Types (from Royer, 1981)

14

Thus, this has implications for comparisons between the original versions of the tests and their updated, computer presented versions used in the UTC-PAB. However, based on the results of Kennedy et al. (1985), it is not expected that these differences will be critical. It is only important that it be kept in mind that it is possible differences do exist.

This UTC-PAB test involves same/different judgements based on the successive presentation of two 8-dot patterns. The patterns are similar to those used by other researchers, including Ichikawa (1981), Klein and Armitage (1979), and Phillips (1974). Since the current task uses successive stimulus presentation, there is a memory loading factor which is present only in one other spatial task in the UTC-PAB. Ichikawa (1981) studied the effects of dot pattern configuration on subjects' estimates of ease of memorization. The results were unequivocal: patterns which were rated as easy to memorize had much higher levels of symmetry than patterns which were rated as difficult to memorize. Thus, it is possible that differential responses based on the perceived symmetry of a given pattern may occur. It may, therefore, be desirable to at least attempt to control some of the more common types of symmetry in order to obtain homogeneous performance within a trial series.

Phillips (1974) evaluated sensory storage and short term visual memory of spatial patterns. He used three different sized matrices; four, six, or eight cells on a side. The density of dots was higher than in the current study; the probability of a cell being filled was 0.5 rather than 0.2. Phillips found that the 4 by 4 matrices had fairly long viable storage times (at least 9 seconds), losing no efficiency over the first 600 msec. In addition, the patterns tended to be resistant to masking or deficits induced by moving or shifting the pattern. In contrast, the larger matrices seemed to be stored in the sensory store and were markedly affected by movement, masking, and storage time. Storage time seemed to be limited to about 100 msec. Thus, it appears that the choice of a 5 by 5 grid with five filled cells for the UTC-PAB version of the test is a viable one, since the dot density is less than in some of the other cited studies. This should result in stimuli that are not highly acceptable to peripheral interference effects (e.g., masking).

Bridgeman and Maye (1983) found that performance was at a chance level when subjects were required to shift fixation from one dot pattern position to another, when trying to locate a single missing dot. Their patterns consisted of 12 dots in a 5 by 5 matrix (making the proportion of filled cells slightly below 0.5) and, for the two separations they used (4 and 2.25 degrees), performance was uniformly poor. Implications for the UTC-PAB version suggest that an overlay of the second stimulus over the first may be the optimal presentation methodology. Another implication is that increasing the number of dots beyond the current five may adversely affect performance.

RELIABILITY

Kennedy et al. (1985) used the Fitts Histograms as a marker test during the development of a microcomputer based repeated measures test battery. They found a test-retest reliability for the task of 0.90. Using the Spearman Prophecy formula, they estimated the reliability of a 3-minute version of the test to be 0.93. The test was administered as a paper and pencil test, which tended to stabilize more slowly than the same test in computer based form. The Fitts Histogram test correlates well with the Klein and Armitage task. In that task, the standard and test stimulus are presented simultaneously rather than successively as in the current experimental test. This makes generalization from that task to the current one less direct, but little data is available otherwise. The primary difference between the Klein and Armitage test and the matrix rotation test is that the latter test loads more heavily on spatial short term memory than the former, which uses simultaneous presentation of stimuli. The Kennedy et al. (1985) study quotes the reliability of the Klein and Armitage (1979) task as 0.93. The reliability of these two tests, and the correlation between them and the current experimental test, implies that the matrix rotation test will also have moderate to high reliability.

VALIDITY

The Fitts Histogram test correlated 0.71 with the Klein and Armitage task (1979) in the Kennedy et al. (1985) study. Previous research has shown

that the Klein and Armitage pattern comparison test loads on spatial factors. Kennedy and his coworkers performed a factor analysis on the tests in their battery (these results should be interpreted with caution since there were only 20 subjects and 11 tests). Three of the tests had both paper and computer versions, three had only computer versions, and two were only administered in paper versions. Of these tests five were predominantly perceptual motor in nature, two were visual, two were spatial, and two were spatial like. They isolated four factors. The Fitts Histograms loaded on the same factors as the Manikin test, code substitution, and the Klein and Armitage task. The most similar test to the current experimental task having validity data available is the Klein and Armitage task. Research by Kennedy et al. (1985) evaluated subject's performance on this task in comparison with standardized tests of intelligence. The Klein and Armitage task correlated 0.57 with the WAIS performance scale, while correlating on 0.05 with the verbal scale. This implies that the task is unrelated to verbal ability. Within the WAIS subtests on the performance scale, the task correlates well with the spatial tests. This pattern of results suggests that the Klein and Armitage test is primarily a spatial task. Since the matrix rotation task is also a dot in matrix type test, it is likely that it also is primarily a spatially loaded task.

SENSITIVITY

Sensitivity to Intrusive Agents and Factors

No research has been completed using the current experimental task to examine the effects of drugs, toxic agents, or environmental stressors. Similar research, however, has been performed on tests which are likely to load on the same spatial factors. The Manikin test has been shown to be sensitive to the effects associated with diving to extreme depth (e.g., 600 meters) (Lewis and Baddeley, 1981; Logie and Baddeley, 1983). The Klein and Armitage (1979) test has been demonstrated to be sensitive to cyclical variations in cerebral hemisphere arousal (Klein and Armitage, 1979). Since it is likely that this test also loads heavily on some of the same spatial factors, it may be conjectured that similar deficits would also occur with the present dot pattern presentation task.

152

TECHNICAL DESCRIPTION

There are 100 pregenerated standard stimuli, one hundred 90 degree right
rotations, one hundred 270 degree right rotations, and 200 nonmatching
stimuli. The standard stimuli are generated with the constraint that at
least one cell will be filled in each row and column. Nonmatching stimuli
will be generated by the displacement of one cell in the matrix, under the
constraints of the generation rule stated above. Responses are made on a
two key response box, with one key labeled "same" and one key labeled
"different."

The stimulus presentations are self paced; the matrices stay on the screen
until the subject presses a key on the response box. Approximately 50 per-
cent of the presentations within a trial will be of identical figures.
Presentations are grouped into 1 minute trials, with a 30-second rest
period between trials. Each subject will receive 20 trials.

Trial Specifications

Each presentation of a standard test pair will consist of the following
steps: (a) the standard stimulus will be presented on the screen until the
subject presses a key on the response box; (b) the test simulus will be
presented and will remain on the screen until the subject makes his same/
different judgement and presses a key; and (c) the next trial will begin.

DATA SPECIFICATIONS

Trial and individual presentation data will be collected. Percent errors
and average correct reaction time will be generated and recorded for each
trial. The mean, standard deviation and range for each 1 minute trial will
be recorded for the error trials and the correct responses separately. In
addition, same/different judgements and 90/270 degree trials will be broken
out as well. Time in viewing the first pattern will also be recorded.

TRAINING REQUIREMENTS

Before the start of the training session, subjects should be read the
instructions to the task. The subjects should receive about 20 minutes of
practice after the instructions; performance on the task should be
approaching asymptote by that time. Presentation during the training
period will be identical to the experimental trials, with the exception
that there will be feedback during the training phase.

To summarize, the training phase for this test should consist of the fol-
lowing steps:

1. Read the instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure
    that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects
    require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run
    over several sessions on this test, one may omit the practice
    trials after the first session.

INSTRUCTIONS TO SUBJECTS

This experiment will examine your ability to mentally rotate one figure to
compare it with another. You will see a 5 by 5 grid, with five of its
cells lighted. You should learn the pattern as quickly and as accurately
as possible, and then press either button on the response box when you are
sure you know the pattern. As soon as you press the key, a new pattern
will be presented. If the new pattern is the same as the old pattern, but
turned 90 degrees to the left or right, press the "same" button on the
response box. If the pattern is not a 90 degree left or right rotation of
the old pattern, press the key on the response box labeled "different." If
you have any questions, please ask the experimenter now.

154

Section 13
MANIKIN TEST (UTC-PAB TEST NO. 12)
(SPATIAL ORIENTATION ROTATION ABILTY)

PURPOSE

The purpose of the Manikin Test is to assess the subject's ability to per-
form rotations and related transformations of a mental image. This ability
is one of the three general subdivisions of spatial ability. Lohman (1979)
has called this ability spatial orientation (SO), which requires mental
movement of the self to view the test stimulus from a new perspective.

DESCRIPTION

The Manikin Test will consist of a series of 64 trials presented to the
subject. On each trial, the subject will see a human figure (the manikin)
displayed on the CRT. The figure will be in one of four orientations: (a)
facing toward the subject; (b) facing away from the subject; (c) right side
up; or (d) upside down. Combinations of all possible pairs of these posi-
tions yields 16 possible orientations for the manikin; a group of these
orientations is a block.

In each hand, the manikin holds a box of a different color (either red or
blue). The manikin stands on a platform that matches the color of a box in
his hand. The subject's task is to indicate the hand (right or left) which
is holding the box that matches the platform color. Responses will be
entered on a response box with two buttons, one labeled "left hand" and one
labeled "right hand." During the 64 training trials (four presentations of
each orientation) the subject will receive feedback; no feedback will be
given during the test trials.

BACKGROUND

Spatial ability is a general term used to describe the human being's facil-
ity for dealing with visually perceived objects and percepts in the envi-
ronment. Lohman (1979) asserts that spatial ability can be broken down

into three separate skills: (a) moving or relocating the mind's eye (or observer's point of view) to a new perspective; (b) rotation and related transformations of mental images; and (c) complex folding and distortions of a mentally imaged object. The Manikin Test seems to tap the rotational transformational aspect of spatial ability.

Spatial transformation has been studied extensively by psychologists (see Cooper and Shepard, 1978 for a review). In fact, Poltrock and Brown (1982) report that the facility subjects exhibit with mental rotation of objects is a good indicator of their spatial ability in general. Many military activities require excellent spatial ability. The most notable of these is piloting aircraft (Egan, 1978), but many enlisted jobs require good spatial ability as well (Carter and Biersner, 1982).

The Manikin Test used in this task appears to involve a mental rotation, the human figure on the CRT is rotated to coincide with the subject's own orientation. After this rotation, the subject makes a response. This pattern of events is supported by the reaction times found by Reader, Benel, and Rahe (1981), who showed that the fastest reaction times were recorded when the manikin was upright and facing away from the observer.

The slowest reaction times were recorded when the manikin was upside down and facing toward the subject. Upon closer examination, it is easy to hypothesize why this is so. Assume that the axes of the manikin are defined as follows: the Y axis is the height, the X axis the width (across the shoulders), and the Z axis the thickness (from front to back). Since the fastest reaction time occurred when the manikin was upright and facing away, it is logical to use that position as the baseline and determine what axial rotations would have to be executed to bring a stimulus into corre- spondence with the orientation of the stimulus with the fastest judge- ment. For the upright and facing orientation, only a Y axis rotation would be necessary. For the upside down and facing away, only a Z axis rotation is needed. But for the upside down and facing orientation, both a Z axis and Y axis rotation are required. A single X axis rotation could also bring the figure into alignment, but the reaction time data are inconsist- ent with that interpretation. If the subject was making such a rotation,

the reaction times would not differ from the other single axis rotations.
Since the reaction times do differ, the two axis rotation seems the more
parsimonious explanation.

Reaction times in a mental rotation task, such as the manikin test, are a
composite of two distinct processes (Cooper and Shepard, 1978). The first
process is the reorientation of the test figure to match the orientation of
the standard maintained in the subject's mind's eye (in this case, upright
and facing away). This is by far the longest of the two processes. The
second component is the time necessary for the actual judgement (i.e., com-
parison of the two stimuli). The amount of time required to make the com-
parison of the two stimuli is usually much less than 1 second. This time,
of course, varies in direct proportion with the complexity of the two
stimuli being compared.

Lohman's (1979) review of many studies from a common theoretical and
statistical standpoint analyzed spatial transformation, as was stated
above, into three distinct abilities. The first, called visualization
(Vz), is the type of mental transformation usually thought of when the term
mental rotation is mentioned. Vz strategies involve the rotation of the
object, while the mind's eye remains stationary. The second type of trans-
formation is called spatial orientation (SO), which involves relocation of
the mind's eye to a new observation position about the stationary stimulus
figure or object. This is the type of mental transformation most likely
required for the Manikin Test. The third general type of spatial trans-
formation is spatial relations (SR). This factor can be best thought of as
the ability to perform any type of spatial transformation quickly. Another
subsidiary factor identified in Lohman's extensive reanalysis of the corre-
lational literature was called the Kinesthetic factor (K). This is the
ability to make left/right judgements, an ability which is likely to play
an important role in Manikin Test performance.

The Manikin Test has several characteristics which make it valuable as a
testing device. Primarily, it is easily learned, since there are only 16
different stimulus orientations. Thus, the subject knows that the stimulus
will appear in only one of those orientations on any given trial. This is

157

different from other tests of spatial ability, which often have a much larger (and in many cases an infinitely large) set of stimuli. The small stimulus set makes it much easier for the subject to focus on the important feature of the stimulus (i.e., the hand holding the box which matches the base).

The second feature of the Manikin Test which makes it experimentally attractive is the fact that, because it is so simple, it takes very little time to administer a large number of presentations. Reader, Benel, and Rahe (1981) administered more than 350 presentations per subject in a 25-minute session. Carter and Woldstad (1985) gave each of their subjects 10 blocks of 80 trials each per day. The Manikin Test was administered as part of a test battery; other tests were given in conjunction with the Manikin.

Finally, the Manikin Test is considered to be more interesting than other tests of spatial transformation, since it involves a human figure. Many tests involve either line drawings of simple or abstract forms, or concrete representations of common objects or views from vehicles. Human beings are intimately familiar with the human form and its configuration; it is assumed that people are more adept at manipulation of such a highly familiar object.

RELIABILITY

The Manikin Test has been in use since the early 1960's (Benson and Gedye, 1963), and thus, has been the object of several reliability evaluations. Reader, Benel, and Rahe (1981) examined the suitability of the Manikin Test for repeated use on the same subject. Their study, using 18 subjects of 3 different age groups and 3 different occupations, found no significant effects for any of these factors over the course of 15 25-minute sessions. In addition, they found no effect for three different types of training schedules.

As a measure of reliability and score stability, the experimenters calculated Pearson product moment correlation coefficients for all pairwise

session mean reaction times. The lowest correlation was .56, but the estimate of common (average) correlation was .84. As a subsidiary measure, each subject was asked to make subjective performance and workload ratings using a simple questionnaire after each session. These ratings did not correlate highly with the reaction times during the session (-.337 and .028 respectively). Each subject was allowed 10 sessions to acquire plateau performance, which was defined as not deviating ± 5 percent from the mean reaction time of the previous two sessions. Plateau performance was reached in an average of 6 sessions (approximately 2300 trials).

Carter and Woldstad (1985) performed a more indepth study of the suitability of the Manikin test for repeated measures, focusing on the validity of using accuracy scores or latency scores as the primary measure for the test. The 20 subjects in this study received 10 blocks of 80 trials per day, over 10 consecutive work days. This represents a 38 percent increase in the number of trials over the Reader et al. (1981) study. Carter and Woldstad's results support the results of the earlier study, with the exception that log latency scores were determined to be better than raw latency data. The log latency scores seem to measure spatial transformation ($r = .38$); the accuracy scores do not ($r = .15$). Thus, log latency scores seem to be the best measure of Manikin Test performance.

Results from the two studies summarized above seem to indicate that the Manikin Test is a useful and accurate test of spatial transformation. It should be noted that the two studies differed in some ways; Reader et al. (1981) used different shapes in the sailor's hands as discriminanda, while Carter and Woldstad (1985) used different colors. The generalized abilities between the two different types of stimuli are not known. The definitive test of reliability and suitability for this test is the Carter and Woldstad effort; the UTC-PAB version shares more methodological similarity with this experiment than with the Reader et al. (1981) version. Further work needs to be performed to determine: (a) if there is a difference between colors and shapes as discriminanda, and (b) whether the performance plateau is the same between the two discriminanda.

VALIDITY

Evaluations of performance on the Manikin Test versus various marker tests indicate that the test appears to measure spatial transformation (Carter and Woldstad, 1985). Correlations on the three marker tests in the study (card rotations, Spatial Apperception Test, Number Comparison) ranged from -.38 to -.49, which was significant at the .05 level. Spatial transformation plays an important role in both the spatial orientation and rotational ability constructs of the subject.

SENSITIVITY

The Manikin Test may be sensitive to some environmental stressors, although the effects of drugs or toxins on Manikin Test performance has not been studied.

The Manikin Test has been applied to several situations involving environmental stress. Lewis and Baddeley (1981) examined the cognitive performance of divers during simulated saturation dives to depths ranging from 300 to 540 meters of seawater. Their results indicated that there were more trials completed on the surface and during decompression than at depth. The differences were small, however, and there were only two divers on each dive. In a related study, Logie and Baddeley (1983) examined cognitive performance decrements during saturation diving with Trimix (helium, oxygen, and nitrogen). Performance on the Manikin Test was relatively unimpaired except at the final depth of 660 meters.

The manikin test has also been applied to the study of acceleration stress on cognitive performance. Lisher and Glaister (1978) studied the effects of $+6_z$ acceleration (the resultant force vector is from head-to-foot) on performance of the manikin test. Lisher and Glaister varied acceleration stress from 1 to 10 $+6_z$ in addition to using three different seat back angles (17, 52, and 67 degrees). Performance on the manikin test was not affected by $+6_z$ acceleration up to and including +6 Gz.

It should be noted that the measure used on the above version of the Manikin Test was number correct (i.e., accuracy), which Carter and Woldstad (1985) have shown to be undesirable. In similar saturation diving studies (O'Reilly, 1977), no significant decrements in spatial orientation ability were found. Thus, it appears that the effect of environmental stressors on Manikin Test performance must, for the time being, remain in question.

TECHNICAL DESCRIPTION

A human facsimile figure will be presented on the CRT, standing with feet apart, arms upraised, and palms up. At the bottom of the screen will be a platform; the ratio of side to base will be approximately 1:4. In each hand, the figure will hold a box, either red or blue. The color of the base will match the color of one of the boxes. The figure will have clearly defined facial features, as well as other detail (clothing detail, et cetera) to insure that it is easy for the subject to judge the figure's position. The figure may appear in one of four orientations of the platform: (a) standing upright and facing toward the subject; (b) standing upright and facing away from the subject; (c) standing upside down and facing the subject; and (d) standing upside down and facing away from the subject.

The figure will remain on the screen for 2 seconds or until the subject makes a response on the response box. There will be two switches on the box, one labeled "left hand," and one labeled "right hand." The stimulus will not be drawn line-by-line on the screen, rather, it will be presented in completed form.

Since there are 16 discrete orientations of the figure, stimuli will be presented in blocks of 16 trials. The test will consist of 6 such blocks, for a total of 96 trials. Data from Reader et al. (1981) indicate that the test, in this form, will take approximately 4 minutes. Each figure will be presented for 2 seconds, with an interstimulus interval of 1 second.

Trial Specifications

Each trial will consist of the following steps: (a) the figure will be
presented in the center of the screen for 2 seconds; (b) at the end of
2 seconds, or as soon as the subject makes a response, the figure will dis-
appear; (c) during training trials, feedback will be presented; (d) the
screen will blank for 1 second; and (e) the next trial will be presented.

DATA SPECIFICATIONS

Response latency will be recorded for each trial. In addition, the sub-
ject's response, the correct answer, and the orientation of the figure will
be recorded for each trial. The summary statistics will include mean and
median response times, their range and variance, and the total number of
correct responses. Trial by trial data will also be available for each
subject.

TRAINING REQUIREMENTS

Before the start of the training trials, subjects should be read the
instructions. After hearing the instructions, each subject should receive
four blocks (64 trials) of practice. The practice trials, unlike the test
trials, will have feedback after each presentation.

To summarize, the training phase for this test should consist of the fol-
lowing steps:

1. Read the instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure
   that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects
   require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

The instructor must be aware of the subject's progress during the practice trials, since the instructions stress speed and accuracy. Additional practice trials may be presented if the experimenter feels the subject is having difficulty with the task or does not understand the instructions.

INSTRUCTIONS TO SUBJECTS

This test examines your spatial ability. The computer will present you with a sailor holding a box in each hand. He will be on another box. The color of the box he is on will match the color of a box that he is holding. The sailor may be facing toward you, away from you, standing up, or standing on his head. Your task is to indicate, by pressing the appropriate button, which hand he is holding the matching box in. You will have only 2 seconds to decide, so you must work as quickly and as accurately as you can. If you have any questions, please ask the experimenter now.

PATTERN COMPARISON (SIMULTANEOUS) (UTC-PAB TEST NO. 13)
(PERCEPTUAL SPEED PATTERN COMPARISON)

PURPOSE

The primary purpose of this self paced pattern comparison test is to assess
the subject's perceptual speed. Perceptual speed is one aspect of general
spatial ability. The test provides information about the subject's ability
to make simultaneous judgements about the similarity of two patterns.

DESCRIPTION

Administration of the test will consist of 60 trials presented to the sub-
ject. On each trial, the subject will see two patterns of eight dots, side
by side on the CRT screen. The pattern on the left is the standard; the
subject's task is to determine if the pattern on the right is identical to
the standard. Responses, entered on a response box, terminate the trial.
If no response is made before the end of a 15-second deadline period, the
trial is terminated automatically. Speed and accuracy feedback will be
given to the subjects during the 10 training trials. No feedback will be
given during the test trials.

BACKGROUND

Pattern perception using figures composed of dots has been studied exten-
sively over the past two decades. One of the most pervasive results of
these dot pattern perception studies is that of goodness of pattern. Good-
ness of pattern is essentially a reflection of the symmetry of the pat-
tern. The effect has been demonstrated in paired associate learning
(Clement, 1967; Glanzer, Taub, and Murphy, 1976), immediate memory
(Attneave, 1955; Home, 1980; Schnore and Partington, 1967), and recognition
and memory search (Checkosky and Whitlock, 1973). The symmetry of the pat-
terns used is important since, according to Howe and Brandau (1983), sym-
metry is processed before form. Symmetry can take several forms. The
first type is called repetition. Repetitions are exact duplicates of a

pattern on both sides of the figure's vertical axis (Figure 12a). A
reflection is a pattern of dots on one side of the vertical axis and the
pattern's mirror image on the other side (Figure 12b). In addition to
these types of symmetry, there are various orders of symmetry. The sim-
plest are first order symmetries, with a single manipulation of the dot
pattern (Figures 12c and 12e). The second order symmetries have four
manipulations of the dot pattern (Figures 12d and 12f). Thus, if one
recognizes that a twelve dot pattern is bilaterally symmetric, the posi-
tions of only six dots need be memorized. The positions of the remaining
six are given. If the pattern's symmetry is of an even higher order, fewer
dot positions will have to be remembered (for example, the subject would
need to learn only three dot positions to be able to reproduce the second
order patterns in Figure 12, once the symmetry had been noted).

This UTC-PAB test involves simultaneous comparison of stimuli. The presen-
tation of figures with symmetry would bias the same/different judgement
reaction times negatively. Thus, the most effective course would be to
exclude from the possible figures either all symmetric or all asymmetric
patterns. The former case is probably the easiest to implement, since
there are fewer symmetrical than asymmetrical patterns.

The bias created by symmetries would be fairly easy to test for, given cer-
tain guidelines. It should be noted that there are about 1800 dot patterns
possible if the 4 by 4 grid is divided into seperate quadrants (i.e., four
different 2 by 2 grids). This number represents the total number of dot
patterns in a 2 by 2 matrix (0 to 4 dots, yielding 16 patterns) in all pos-
sible combinations of four 2 by 2 matrices. Of these possible 1800 4 by 4
dot matrices, only about 400 have eight dots, the number required for this
experimental configuration. The 400 patterns are created from the total
possible without replacement, (a given 2 by 2 matrix can only occur once
within any specific 4 by 4 matrix). In addition, rearrangements of 2 by 2
matrices do not repeat either (i.e., if one possible pattern is ABCD, the
pattern CBDA is not valid since it is merely a repetition of the first).
Since the smaller matrices do not repeat, the possibility of apparently
symmetrical patterns is greatly lessened. In addition, the 400 standard

```
           |                              |
        ** | **                        ** | **
         * | *                          * | *
        ** | **                        ** | **
         * | *                          * | *
```

a) Repetition   b) Reflection

```
           |                              |
        ** | **                         * | *
         * | *                         ** | **
        ** | **                        ** | **
         * | *                          * | *
           |
```

c) First Order Reflection   d) Second Order Reflection

```
           |                              |
        ** | **                         * | *
         * | *                         ** | **
        ** | **                         * | *
         * | *                         ** | **
```

e) First Order Repetition   f) Second Order Repetition

Figure 12.   Examples of Symmetry

166

patterns make it possible to generate the "different" stimuli in a standard fashion as well. Since the stimuli can be the same across subjects, the ability to generalize will be enhanced. This enhanced generalized ability would not be present in an experiment with stimuli created randomly for each subject.

The choice of a 4 by 4 grid for this test is guided by the work of Phillips (1974). His study looked at short term visual memory in a same/different judgement task using 16, 36, and 64 element matrices. In the task, the subjects were required to decide if two patterns were the same; patterns were made different by removing a single dot from the matrix. Note that this is quite similar to the displacement of a dot in the current experimental paradigm. In addition, varying delays were introduced between the offset of the standard stimulus and the onset of the test stimulus. The patterns were often quite complex, since a matrix cell had a 50 percent chance of being filled. Phillips' results suggested that for the 4 by 4 cell matrices, there was some decline in performance over the first 600 msec of the delay period, though the subjects performance stabilized over the longest delay used (9 seconds). The smallest grid also showed strong resistance to masking and stimulus movement. On the contrary, the larger grids proved to be highly susceptible to both movement and masking of any kind. The isomorphism between the smallest grid size and the grid in this study, and the general experimental paradigm, make it safe to assume that the UTC-PAB version will be both easily implemented for administration and easily learned and performed by the subjects.

Klein and Armitage (1979) used same/different judgements of dot patterns in a study of cyclical variations in cognitive style. In their study, subjects were shown a dot pattern on the screen for a short duration, which was then removed from view. A second pattern was then presented, and the subject was required to decide whether the second pattern was the same as the first. It is difficult to surmise exactly how their task compared to the current one, since the brief format of the article left little room for details concerning stimulus construction. However, the nature of their study and their results imply that the test did, in fact, measure some facet of spatial ability. Specifically, they were attempting to find

167

regular variations in hemispheric activation. To find these variations, they administered a left hemisphere task (a semantic judgement task) and a right hemisphere task (the dot pattern task) at regular, short intervals throughout the day. Their analysis concentrated on changes in the performance of the tasks as a function of time of day. Their results implied that there was a cyclical (and nearly sinusoidal) variation in test performance on both tasks. Moreover, the cycles on the two tests were 180 degrees out of phase with each other, strongly suggesting that there is a regular and periodic change in hemispheric activation. The fact that the dot pattern test was different from the verbal task, thus, implies that the Klein and Armitage task does assess some aspect of spatial ability.

One major difference between these other uses of same/different judgements of dot patterns and the current experimental paradigm is the relative speed allotted to the subject for their response. In the Klein and Armitage (1979) task, the subjects were told to complete as many test items as possible in the available time. In Phillips' (1974) study, the time to respond was measured, but was relatively open ended. In the current task, however, the subjects must make their judgement in a very short, fixed time interval.

In Lohman's (1979) reanalysis of the correlational literature, the factor of response speed (in the sense of the time window within which the subject must respond) played an important role. He found that, given the same test, introduction of time constraints to a test drastically changed the spatial factor being measured by the test. Only three general spatial factors (or abilities) emerged from the review. The highest level factor is called visualization (Vz). It appears in tasks requiring the mental reorientation of a highly complex form or object. Vz tasks can usually be recognized by relatively slow responses. The second factor is spatial orientation (SO). Tests assessing this factor involve the ability to imagine how a stimulus will appear from a different perspective. This type of task involves a mental reorientation of one's self, rather than the object in the problem. The final spatial factor is called spatial relations (SR). These types of tests are the most highly speeded of the spatial ability tasks. Again, mental rotation seems to be the common

168

element, but the primary factor seems to be the ability to solve spatial problems quickly, by whatever means.

The various facets of spatial ability can te arranged in a hierarchy. One of the most useful graphic representations was presented in Figure 10. The factors can be characterized along two dimensions: speed/power and simplicity/complexity. The more powerful an ability, tne higher its position in the factor hierarchy. However, a higher position in the hierarchy also guarantees slower performance, since the tasks are more complex. There are four other spatial factors at the bottom of the hierarchy which have not been discussed up to this point, but deserve mention: Closure speed (Cs), the speed of matching incomplete or distorted stimuli with representations in long term memory; Kinesthetic (K), the speed of making left/right decisions; Visual memory (M), the ability to maintain stimuli in short term memory; and Perceptual speed (Ps), the speed of matching stimuli. The reader will note that all of these factors might play a part in the test under consideration here. The primary loading for this test, however, would probably be on the Cs and Ps factors. These two factors are exactly the constructs that this test was chosen to measure. Note that these factors are all at the lowest level of Lohman's hierarchy; this implies that the test might have many factors in common with all of the higher level spatial ability constructs (most notably Vz and SO).

RELIABILITY

Kennedy et al. (1985) in their evaluation of a number of tests for a portable microcomputer repeated measures testing system, quote the reliability of the Klein and Armitage (1979) task as .93. That task is the same as the current one, in that presentation is simultaneous rather than successive, so the two tasks are similar enough that some conjecture may be drawn as to the reliability of the test.

VALIDITY

The Pattern Comparison task used by Klein and Armitage (1979) is similar to the current experimental task. Research by Kennedy et al. (1985) has

evaluated performance on this task in comparison to performance on standardized tests of intelligence. The Klein and Armitage task had a correlation of .57 with the performance scale of the Wechsler Adult Intelligence Scale (WAIS); the correlation with the verbal scale of the same test was only .05. This implies that the current experimental task is not a verbal one.

Within the performance subtests, the pattern comparison task correlated most highly with the Digit Symbol Substitution test (.71), followed by the Block Design test (.59), Picture Arrangement (.29), and Object Assembly (.27). All of these tests involve visual scanning of a standard and mental and physical manipulation of various component parts to construct a duplicate of the standard. These tests are all spatial in nature, and the correlations they show with the Pattern Comparison task suggest that it, too, is a spatial task.

SENSITIVITY

There is little available data on the effects of drugs, toxic agents, or environmental stressors on the specific test addressed in this manual. However, there are some indications of effects on other tests which load on some of the same spatial factors. The Manikin Test has been shown to load on the spatial transformation factor (most probably SO) (Carter and Woldstad, 1985); performance on that test shows a severe decrement when it is administered to divers at extreme depth (Lewis and Baddeley, 1981; Logie and Baddeley, 1983). Since it is likely that the present test also loads heavily on some spatial factors, it may be assumed that such a deficit would also occur under the same environmental stress for this dot pattern presentation task.

TECHNICAL DESCRIPTION

The patterns will be generated in a random fashion on a 4 by 4 grid. After the pattern is generated, a test for repeated and reflected figures will be conducted. Any such figures will be discarded. After the first pattern

has been generated, a random determination of whether to plot the same or a different pattern is made.

Once both patterns have been generated, they will be displayed on a light blue background. Each pattern will be enclosed in a box with a dark blue border, the dots will be white.

The only valid keys will be the two marked "same" and "different" on the response box. Depressing any other key will have no effect. Key presses will not be echoed to the screen.

Trial Specifications

Trials will proceed in the following fashion: (a) a pair of patterns will be presented on the screen; (b) the subject presses the key labeled "same" or "different" according to his judgement before the time limit elapses; (c) the screen will clear for 500 msec; (d) during practice trials, after an incorrect response, the screen will display the correct response for 5 seconds at which time the same trial will be repeated; (e) a new trial will be presented.

DATA SPECIFICATIONS

The program will generate and record two dependent measures for each trial: (a) RT: The reaction time of the subject's same/different judgement, measured from the initial presentation of the two patterns. (b) Response Code: The response code indicates the response made (e.g., correct, incorrect, or terminated by the deadline). In addition, trial type (e.g., same or different) will be recorded for each trial.

Summary statistics will be provided for the same trials, different trials, and overall trials. Statistics will include the mean and median response latency, the range and the variance of the latencies, and the total number of correct trials. Data may be examined on a trial by trial basis, with each trial's response latency, response accuracy, and trial type

displayed. Hardcopy of the trial by trial and summary statistics will be available.

TRAINING REQUIREMENTS

Initially, subjects should be given the instructions that follow. After the instructions, the subjects should be presented with at least 10 practice trials. Presentation during the practice trials will be identical to the test trials. However, during practice, the subject will be given feedback after each incorrect trial. After the feedback, the same trial will be repeated.

The test administrator should be acutely aware of the subject's performance during the practice session. It is important to be sure that the subject is following the instructions; they should be responding as quickly and as accurately as possible.

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

This test examines your ability to compare two patterns simultaneously. The computer will present two patterns of dots to you, side by side on the

172

screen. You must decide, as quickly and accurately as possible, if the two patterns are the same. You will indicate your answer by pressing the button labeled "same" on the response box in front of you if the patterns on the screen are identical, or "different" if the patterns are different. Once you press a button, the patterns will disappear, so it is important that you know your answer before you press either button. If you do not answer in 15 seconds, the patterns will disappear and new ones will be displayed.

Section 15
PATTERN COMPARISON (SUCCESSIVE) (UTC-PAB TEST NO. 14)
(PERCEPTUAL SPEED SHORT TERM SPATIAL MEMORY)

PURPOSE

The primary purpose of this task is to examine the subject's short term
spatial memory and perceptual speed. The test is diagnostic of spatial
memory, since the subject must maintain the standard in memory while the
comparison with the test pattern is being made.

DESCRIPTION

The test will be administered as a series of 60 trials presented to the
subject. Each trial will proceed in the following fashion: The standard
pattern will be presented for 1.5 seconds. At the end of that period, the
screen will clear for 3.5 seconds, at which time the second (or test) pat-
tern will be presented. The test pattern will remain on the display until
the deadline period expires (15 seconds) or the subject makes a response.
The subject's task is to determine whether the two dot patterns are the
same or different.

During the training phase, the subject will respond to 10 trials. Response
speed and accuracy feedback will be provided to the subject after each of
the training trials. Feedback will not be presented during the test
trials.

BACKGROUND

Over the years, an extensive body of research into spatial perception has
developed. For the most part, study of the various abilities man has
evolved to manipulate spatial information has been examined in isolation,
with the individual researcher evaluating a specific ability within a spe-
cific theoretical framework. Various hierarchies and heterarchies of spa-
tial ability have been developed, but for the most part the resultant
frameworks have been little more than weakly supported hypotheses. Then,

174

in the early years of the twentieth century, researchers began using factor analytic techniques. This correlational analyses allowed the researcher to compare many tests at a time and, in addition, determine how closely various subsets of a test battery were interrelated. When these statistical methods were applied to tests purported to measure spatial ability, it was found that certain types of tests clustered together (which is to say they seemed to measure the same factor) while others were separated from the cluster (or measured different abilities). These analyses implied that spatial ability could be characterized by several different skills.

By the mid 1970s, a great deal of factor analytic work had been performed, much of it with the intent of delineating the extent and nature of spatial ability. This body of research, however, was diminished in usefulness by the constant plague of the researcher: different procedures, different measures, different numbers of subjects, different program intents, and different theoretical frameworks. Comparison and evaluation of different studies were and are quite difficult. Lohman (1979) attempted to clear up some of the difficulties through a two step process: (a) analyze the data from the studies using the same procedure throughout; and (b) interpret the results from a common theoretical perspective. Lohman's results were both interesting and valuable. Only three general spatial factors (or abilities) emerged from this review. The highest level factor is called visualization (Vz). It appears in tasks requiring the mental reorientation of a highly complex form or object. Vz tasks can usually be recognized by the relatively slow nature of their performance. The second factor is spatial orientation (SO). Tests assessing this factor involve the ability to imagine how a stimulus will appear from a different perspective. This type of task involves a mental reorientation of one's self, rather than the object in the problem. The Manikin Test (see UTC-PAB Test No. 12) probably falls into the category of an SO test. The final spatial factor is called spatial relations (SR). This factor can be thought of as the ability to perform any type of spatial transformation quickly. Again, mental rotation seems to be the common element, but primarily the factor seems to represent the ability to solve spatial problems quickly by whatever means.

175

Within these three general types of spatial abilities are two types of spatial transformation: mental movement and mental construction. Mental movement can be thought of as rotation, translation, folding, movement of, or movement around a mental image of a stimulus. Mental construction involves either the physical assembly of a stimulus from a mental representation (for example, by drawing or building a facsimile of the stimulus), or mental combination (mentally joining together separate images to form a larger, more complex image). This UTC-PAB test most likely loads on the mental movement aspect of spatial transformation.

The various facets of spatial ability can be arranged in a hierarchy. One of the most useful graphic representations was presented in Figure 10. The factors can be characterized along two dimensions: speed/power and simplicity/complexity. The more powerful an ability the higher its position in the factor hierarchy. However, a higher position in the hierarchy also guarantees slower performance, since the tasks are more complex. There are four other spatial factors at the bottom of the hierarchy which have not been discussed, but deserve mention: Closure speed (Cs), the speed of matching incomplete or distorted stimuli with representations in long term memory; Kinesthetic (K), the speed of making left/right decisions; Visual memory (M), the ability to maintain stimuli is short term memory; and Perceptual speed (Ps), the speed of matching stimuli. The reader will note that all of these factors might play a part in the test under consideration here.

Contrary to the views of other researchers, Lohman asserts that: "Mental rotation, while an interesting and special type of mental transformation, is not the most important determinant of spatial ability. Rather, the crucial components of spatial thinking may be the ability to generate a mental image, perform various transformations on it, and remember the changes in the image as the transformations are made. This ability to update the image may imply resistance to interference, both internally and externally generated. Further, it implies that one of the crucial features of individual differences in spatial ability may lie not in the vividness of the image, but in the control the imager can exercise over the image" (1979, page 116).

Currently, one of the major problems in spatial perception research is the fact that little control is exercised over the subject's choice of problem solving strategies. With a small number of subjects, it is not difficult to evaluate each response to insure that the desired strategy is being used (i.e., for a Vz task, reorienting the imaginary object rather than the self). However, this problem becomes much greater as the number of subjects increase. With tests such as those in the UTC-PAB, it is safe to assume that the tests will be administered to a large number of subjects (more than 100); thus, it is important to consider the disparities induced in the data by the use of different strategies. Research has shown that more often than not, subjects use different strategies to solve the same test. Within a test, the number of distinct strategies will increase as item difficulty and complexity increase. There will be a concomitant decrease in response speed as complexity increases. However, even on the most simple speed tests, subjects still can be relied upon to use different strategies. Tests which the researcher intends to be solved using one strategy are often solved using another. For example, early researchers had great difficulty seperating Vz and SO tests. It was not until they realized that SO tests were often solved using Vz strategies that the differentiation became more reliable. And finally, mental manipulation is often discarded in favor of more analytic methods as complexity and difficulty increase (i.e., the subjects count angles or note distictive features instead of using mental transformation to solve the problem).

It is obvious that various spatial abilities (probably three) are present and available to the subject. However, caution must be used in any test of spatial ability. Tests are solved in different ways by different subjects. Their solution strategies change as a function of myriad factors, including practice and item difficulty. Further, most factors represent individual differences in speed of solving particular types of problems, not general problem solving skills or abilities. Finally, the process of adapting a test to an experimental task may drastically alter the nature of the test. An experimental task will rarely tap exactly the same mental processes as the source test.

177

The UTC-PAB test involves same/different judgements based on the successive presentation of two eight dot patterns. The patterns are similar to those used by other researchers, including Ichikawa (1981), Klein and Armitage (1979), and Phillips (1974). The differences are worth noting however. Ichikawa was studying ease of dot pattern memorization. He used eight dot patterns in a 4 by 4 matrix, and seven dot patterns in a 3 by 5 matrix. Through the use of a complicated metric, various types and levels of symmetry for each dot pattern were computed. These values were then applied (through multiple regression) to the results of a subjective rating of each pattern on a nine point ease of memorization scale. The results were unequivocal: patterns which were rated as easy to memorize had much higher levels of symmetry than patterns which were rated as difficult to memorize. Implications for this study include possible differential responses based on the perceived symmetry of a given pattern. Thus, it may be desirable to at least attempt to control for some of the more common types of symmetry.

Klein and Armitage (1979) used seven dot patterns in a simultaneous pattern comparison task. It is unclear in what size matrix the dot pattern was embedded. Their study was intended to evaluate performance differences as a function of biological rhythms. These rhythms involved an alternation in the relative efficiency or activation of the two cerebral hemispheres. Klein and Armitage reasoned that, since the two hemispheres reflect different cognitive functions, frequent administration of two tests targeted for each hemisphere should demonstrate cyclical changes in performance. Their study showed just such a cycle, on the order of 90 minutes in length.

Phillips (1974) evaluated sensory storage and short term visual memory. He used three different sized matrices, four, six, or eight cells on a side. The density of dots was higher than in the other studies mentioned; the probability of a cell being filled was 0.5. He found that the 4 by 4 matrices had fairly long viable storage times (at least 9 seconds), losing no efficiency over the first 600 msec. In addition, the patterns tended to be quite resistant to masking or deficits induced by moving or shifting the pattern. In contrast, the larger matrices seemed to be stored in the sensory store and were markedly affected by movement, masking, and storage

178

time. Storage time seemed to be limited to about 100 msec. Thus, it
appears that the choice of a 4 by 4 grid for the current study is the most
viable one, based on the successive comparison paradigm.

Bridgeman and Mayer (1983) found that performance was at a chance level
when subjects were required to shift fixation from one dot pattern position
to another when trying to locate a single missing dot. Their patterns con-
sisted of 12 dots in a 5 by 5 matrix and, for the two separations they used
(4 and 2.25 degrees) performance was uniformly poor. Implications for the
UTC-PAB version suggest that an overlay of the second stimulus over the
first may be the optimal presentation methodology.

RELIABILITY

Kennedy, Wilkes, Lane, and Hanick (1985) in their evaluation of a number of
tests for a portable microcomputer repeated measures testing system quoted
the reliability of the Klein and Armitage (1979) task as .93. That task
differs from the current one in that presentation is simultaneous rather
than successive, but the two tasks are similar enough that some conjecture
may be drawn as to the reliability of the test successive presentations.

VALIDITY

The Pattern Comparison task used by Klein and Armitage (1979) is similar to
the current experimental task. Research by Kennedy, Dunlap, Jones, Lane,
and Wilkes (1985) has evaluated performance on this task in comparison to
performance on standardized tests of intelligence. The Klein and Armitage
task had a correlation of .57 with the performance scale of the Wechsler
Adult Intelligence Scale (WAIS); the correlation with the verbal scale of
the same test was only .05. This implies that the current experimental
task is not a verbal one.

Within the performance subtests, the pattern comparison task correlated
most highly with the Digit Symbol Substitution test (.71), followed by the
Block Design test (.59), Picture Arrangement (.29), and Object Assembly
(.27). All of these tests involve visual scanning of a standard, and

179

mental and physical manipulation of various component parts to construct a duplicate of the standard. These tests are all spatial in nature, and the positive correlations they show with the Pattern Comparison task suggest that it, too, is a spatial task.

SENSITIVITY

There is little available data on the effects of drugs, toxic agents, or environmental stressors on the specific test addressed in this manual. However, there are some indications of effects on other tests which load on some of the same spatial factors. The Manikin Test has been shown to load on the spatial transformation factor (most probably SO) (Carter and Woldstad, 1985); performance on that test shows a severe decrement when it is administered to divers at extreme depth (Lewis and Baddeley, 1981; Logie and Baddeley, 1983). Since it is likely that this test also loads heavily on some spatial factors, it may be predicted that such a deficit would also occur under the same environmental stress.

TECHNICAL DESCRIPTION

The eight dot patterns used in the study will be generated on a 4 by 4 grid. After the first pattern is generated, a test for repeated and reflected patterns will be carried out, and any such figures found will be discarded prior to display. After generation, a random determination to display the same figure or a different one will be made. If the figure is to be different, three dots will be displaced in the original, using the noticeable difference algorithm developed by Irons (1984). This will become the figure labeled "different."

At this time, the standard pattern will be presented on the screen, centered on a light blue background, and enclosed within a dark blue box. The standard will be presented for 1.5 seconds. At the end of this period, the screen will blank for 3.5 seconds, at which time the test stimulus will be presented. The presentation of the test stimulus will last 15 seconds.

180

Only two keys (on the response box) will be valid; they will be labeled "same" and "different." Pressing any other key will have no effect. The computer keyboard keys will be ignored.

Trial Specifications

Each trial will take place in the following sequence: (a) a pattern will appear, centered on the screen, for 1.5 seconds; (b) the screen will clear for 3.5 seconds; (c) the test pattern will appear for 15 seconds, or until the subject enters a same/different judgement response; (d) during practice trials, feedback will be provided to the subject for 5 seconds; (e) the screen will clear and the next trial will begin.

DATA SPECIFICATIONS

The program will generate two measures for each trial: (a) RT: Reaction time of the subject's same/different judgement, measured from the initial presentation of the test stimulus. (b) Response Code: The classification of the subject's response (e.g., incorrect, correct, or terminated by the deadline).

In addition, trial type (same or different) will be recorded for each trial.

Summary statistics which will be computed will include total elapsed time for the task, number and percent correct, and number of trials terminated by the deadline. Reaction time means and standard deviations will be computed for each trial, broken out by all trials, correct trials, and incorrect trials. Trials terminated by the deadline will not be included in any calculations.

TRAINING REQUIREMENTS

Before beginning the experimental run, subjects should be read the instructions. After hearing the instructions, the subjects should be given at least 10 practice trials. Presentation during the practice trials will be

identical to the experimental trials, with the exception that the subject
will receive feedback only during the training trials. Feedback will be
given only after an incorrect response during the training; the missed
trial will be repeated.

The person administering the test should closely monitor the subject's
performance during the course of the training. The experimenter should be
sure that the subject understands both the instructions and the task, and
is performing at an acceptable level. The instructions stress fast and
accurate response; the subject's performance should not sacrifice one
aspect for the other.

To summarize, the training phase for this test should consist of the
following steps:

1. Read instructions to tne subjects.

2. Run practice trials and evaluate subjects' performance to ensure that
    the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects' require
    additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over
    several sessions on this test, one may omit the practice trials after
    the first session.

INSTRUCTIONS TO SUBJECTS

This test examines your ability to compare two patterns, presented one
after the other. The computer will present two patterns of dots to you.
You should try hard to remember the first pattern. After a short time on
the screen, it will be erased, and a second pattern will be displayed. You
must decide if the second pattern is the same as or different from the
first. If you think the second pattern is different from the first, press
the key on the response box labeled "DIFFERENT." If you think the two

patterns are the same, press the key labeled "SAME." It is very important to give your answer as quickly as you can without making mistakes. As soon as you give your answer, the screen will clear again and a new pair of patterns will be presented. Before we begin, you will be given some practice runs. The experimenter will tell you when the test begins. If you have any questions, please ask the experimenter now.

Section 16
VISUAL SCANNING TASK (UTC-PAB TEST NO. 15)
(PERCEPTUAL SPEED)

PURPOSE

This task is a modification of Neisser's (1963) letter search task which requires subjects to search for and detect a target embedded in nontarget items. This test is diagnostic of a subject's ability to perform rapid visual pattern discrimination.

DESCRIPTION

The UTC-PAB visual scanning task can be presented in one of two alternative versions. Both procedures require that the subjects visually scan a matrix of letters (25 rows by 5 columns) in normal reading order (left to right, top to bottom) in order to detect a prespecified target letter (e.g., "K") embedded in the matrix. In the lite pen version, once the target letter is detected, the subject is required to identify the exact location of the target using a lite pen. In the keyboard version, the subject identifies the row of the matrix in which the target is embedded via a keypad or keyboard.

BACKGROUND

The visual scanning procedure was developed by Neisser (1963) in order to provide information about the depth, breadth, and flexibility of tne cognitive processes involved in recognizing printed letters. The test is based on the theory that the process of recognition is hierarchically organized. That is, before a subject decides that the letter Z, for example, is present in the stimulus display, prior decisions must be made about features of the stimuli such as parallel lines and angles. These decisions, in turn, are then based on processes of a still lower order (e.g., "feature" detectors in the visual system). According to the theory, processing times would be expected to depend on the depth of hierarchy required by the task. If, however, several operations are at the same

184

level in the hierarchy, the subject may be able to execute them simultaneously (in parallel).

Neisser's (1963) original study consisted of several variations of the basic procedure. In the first experiments the identity of the target letter (Q versus Z), the number of columns in the matrix (two versus six), and the presence or absence of the target letter were all varied. The study involved the additional manipulation of such variables as the horizontal spacing of the rows; the context in which the target was embedded (angular letters such as W, X, and Y, or round letters such as G, O, and U); the number of days of practice; and the number of targets searched for.

The results of the study were as follows: (a) it takes longer to detect the absence of a letter than its presence, (b) subjects can look for either of two letters as rapidly as for one alone, (c) the more columns in the display, the longer it takes to detect the presence or absence of a target, (d) context plays an important role in feature detection (e.g., it is easier to detect a round nontarget letter in a context of angular nontarget letters than in round target letters), (e) reaction times decrease with practice, and (f) with enough practice subjects searched as quickly for four targets as for one target.

The following conclusions were offered with regard to the cognitive processes involved in the identification of printed letters: (1) At simple levels, several distinct processes of recognition can function simultaneously (i.e., in parallel) in the analysis of a single stimulus configuration. However, (2) parallel processing does not appear to be evident in the analysis of "spatially distinct" parts of the input, even after extended practice. (3) The nature of the search process is dependent upon the nature of the context in which the target letters are embedded.

Many researchers have subsequently investigated the scanning task in order to more precisely define the cognitive processes involved. The majority of this literature centers around identifying those factors that are most necessary for parallel processing of letters (i.e., when scanning times remain constant as the number of targets searched for increases). Four such

185

factors have been determined to separate parallel processing (visual scanning functions) from serial processing (item recognition functions). These factors are: (1) the amount of practice with the search task, (2) the nesting of target sets, (3) the analyses of speed and errors, and (4) the context of the target. Representative studies for each of the above factors will now be presented.

The Effects of Practice

Neisser, Novick, and Lazar (1963) presented a study in which subjects searched for targets of sizes 1 to 10 letters. The results showed that by the twelfth day of practice, reaction times were the same for all number of targets searched for. That is, by the twelfth day, subjects could search for 10 targets as quickly as they could search for one. This supports a parallel processing model. Error rate was 20 percent in the experiment. In contrast to this, Kaplan and Carvellas (1965) tested the hypothesis that scanning time for just learned targets increases in proportion to the number of targets being searched for. Their results showed that, for target sets of one to five, scanning time was proportional to the number of targets searched for supporting a serial processing model with unpracticed subjects.

Graboi (1971) investigated the effect of specific versus nonspecific practice on scanning speeds. With specific practice (retaining the same stimulus items for all set size conditions) visual scanning rates remain constant over larger set sizes supporting Neisser's parallel processing model. However, when target items differ for every set size condition (nonspecific practice), the search rate increases with set size supporting a serial model of processing (Sternberg, 1969). To explain these results, Graboi argued that the effect of practice might be to develop selectively those feature analyzers relevant to the specific target set, reducing the cues needed to recognize the target and distinguish it from nontarget items. Since the decision process needs to reckon with fewer features, categorization time per item decreases. As a result, the dependence of scan time on memory set size becomes reduced.

Nested Target Sets

Closely related to the effects of nonspecific and specific practice are the effects of nested target sets. Nested target sets occur when each target set contains all the items also contained in smaller sets, and target sets are constant throughout the experiment, as was the case in Neisser's original study.

Kristofferson, Groen, and Kristofferson (1973) listed three conditions which differentiate visual search functions (Neisser, 1963) from item recognition functions (Sternberg, 1969). That is, there are three conditions necessary for search times to be independent of set size: (1) Error rate is high--20 percent, (2) constant and nested targets must be used, and (3) there must be response consistency (always responding in the same manner). The present experiment maintained low error rate, nonnested targets, and response inconsistency in collecting visual search data. Results showed that search times increased in a linear fashion with increases in target set size. Thus, it was concluded that the effect of set size and the effect of practice on the set size effect as determined from visual search performance is qualitatively very similar to the effect of set size and the effect of practice on set size as determined from item recognition performance.

Another study using nonnested target sets was reported by Gould and Carn (1973). Subjects searched for one, five, or 10 targets, any one of which had to occur once, twice, or four times in the array. Different subsets were selected for targets from the 10 target condition every day. The remaining items not chosen in the subset ("nontarget targets") were also presented in the matrix. Results showed search times decreased over a period of 30 days, however, they increased as a function of the set size. A new finding was that subjects required more time and made more errors when searching for five targets than when searching for 10 targets. Two explanations were proposed. First, subjects' verbal reports indicated that the presence of nontarget targets in the array caused an interference effect. When subjects fixated on a nontarget element they had to stop and think whether it was really a target or not. The other explanation

involves the nesting of the targets. Although the items in each target set size remained the same for 42 consecutive trials each day, the items in 1- and 5-item sets changed from day to day, whereas items in the 10-item set remained the same every day.

Holmes et al. (1978) provide an excellent review of the situational factors involved in the visual search paradigm. Holmes et al. (1978) used non-nested target sets that varied from trial to trial. Stimuli consisted of geometric forms which were used to eliminate verbal rehearsal. The results do not provide any support for the existence of parallel processing. Performance throughout the experiment steadily declined as the number of items in the target set increased. These findings provide further support to those of Gould and Carn (1973) and Kristofferson et al. (1973) and suggest that parallel processing cannot be observed unless nested target sets are employed. According to Gould and Carn (1973) the need to learn new target sets on every trial is a difficult task. If nested target sets are used, it is probably much simpler to learn the "master" set (of which all other sets are subsets) and use this master set on all trials, resulting in data which resemble parallel processing.

Speed and Error Analyses

Another factor affecting visual scanning times is the subject's allocation of speed versus accuracy in the task. In Neisser's original study, subjects were told to scan the array as fast as possible. With speed being stressed, the error rate was 20 percent.

Cohen and Pew (1970) replicated Neisser's study in every respect except that accuracy was stressed as opposed to speed. Search times were longer for all target set sizes. After 15 days of practice, search times were not constant for all target sizes, although with succeeding days the differences in time per element associated with the number of possible targets became markedly less.

Wattenbarger (1968) used a speed group and an accuracy group to test the effect of different instructions on scanning speeds. The speed stressed

group had an error rate of 15 percent while the accuracy stressed group had an error rate of 7 percent. Wattenbarger states that this difference in error levels indicates that the verbal instructions were adhered to. Search rates obtained by the accuracy group were slower than for the speed group and performance continued to improve with practice for both groups. It was concluded that a lenient accuracy criterion is necessary, as well as practice, to produce parallel information processing in a visual search task.

Kristofferson (1972) criticized the results of Neisser because the error analysis was based only on the frequency of occurrence of false-negative errors (failure to find the target), and false-positive errors (finding an incorrect target) were not examined. Kristofferson replicated Neisser's original study to allow for measurement of both types of errors. False-positive errors could be identified since the subjects responded by marking the position of the target using a lite pen. Results showed that there were significant differences in scanning times as a function of set size over the final eight days of the experiment. Both types of errors were low. It was concluded that parallel processing and highly accurate performance on the search task are incompatible.

Context Effects

The context or background in which target letters are scanned have also been shown to effect search times. Gould and Carn (1973) varied the background in which targets and nontargets appeared. A complex background consisted of stimulus items located between columns of "percent" symbols. The effect of the complex backgrounds was to add a constant of about 1 second to all search times.

Context of nontarget items has been manipulated by Tone (1981). Subjects searched quickly for all possible target "Zs" in the array. When a target was found, the letter was crossed out with a pen. Either one, two, or four nontarget letters were interposed between targets to make up a 6 by 22 array. Three types of interposed letters were used, angular, round, and both or mixed. Round letters consisted of B, C, D, G, O, and Q while the

angular letters were A, K, M, N, V, and W. The results confirmed expectations. Visual scanning time decreased significantly as the span between targets expanded from one to four interposed letters. Scanning time significantly increased, however, as the visual difficulty of the tasks increased from checking Zs among round letters to checking Zs among angular letters. There was no significant interaction.

In summary, results from visual scanning experiments have shown mixed results. Although all scanning times have been shown to decrease with practice, multiple target set scanning times do not always equal single target scanning times with extensive practice. It seems that for parallel processing of letters to occur, speed at the expense of accuracy must be stressed, and target sets must be nested.

The UTC-PAB version of the visual scanning test involves using only single targets. Subsequently, findings involving nested target sets are not pertinent to this version. However, the effects of practice and speed stress should be controlled. Finally, since varying the context appears to be a good discriminator of scanning times, it should be considered for inclusion as a possible independent variable in this version of the test.

RELIABILITY

Carter and Krause (1983) tested the reliability of both slope scores and response time measures of Neisser's visual scanning task. Twenty three subjects were tested in the experiment in which subjects scanned lists for one of one, two, or four prespecified targets. The probability of finding a target was .50. Subjects were allowed 20 seconds to search in the one target condition and 30 seconds for the other two conditions. The test was repeated in the same order on each of 15 successive weekdays.

The intertrial correlations of the response times for the 15 days for both one and four targets were computed. The average one target correlation is .58 and the average four target correlation is .44. The slope intertrial correlation is .30. Thus, the reliability of slope score was poor compared with the reliability of the RT scores for this letter search task. The

authors, therefore, concluded that response times are more reliable measures of visual scanning performance than are slope scores.

It is important to note that test-retest reliabilities (intercorrelations) obtained on this task risk being confounded by practice effects. Since response times have been found to decrease as a function of days of practice and as a function of target set size (Neisser, Novick, and Lazar, 1963), intertrial correlation values will be biased or contaminated by these effects. Results show that response times for multiple target sets match that of a single target by about the twelfth day of practice, but response times for all target sizes continue to decrease through 30 days of trials.

VALIDITY

The original purpose of the Neisser visual search task was to provide preliminary information about the depth, breadth, and flexibility of the processes involved in recognizing printed letters. If a subject scans at the fastest rate consistent with relatively error free performance, this rate should be limited only by the speed with which he can analyze the items for the presence of a particular letter. Although this task measured the speed of the perceptual (scanning) proces^, it was not the speed alone that the task was designed to measure.

In multiple target searches scanning times vary greatly across experiments. In some studies the scanning times do not increase with more targets, supporting parallel processing (Neisser, Novick, and Lazar, 1963), while in other studies response times increase with added number of targets resembling item recognition functions (Kristofferson, Groen, and Kristofferson, 1973). It is possible that with multiple target comparisons, the task also requires certain memory comparison processing times in addition to perceptual processing times.

In summary, the UTC-PAB version of the visual scanning task, searching for only one target, does seem to tap a person's speed for making rapid visual discriminations.

SENSITIVITY

There is very little literature showing the use of the Neisser visual
search paradigm as a test of human performance in different settings. The
Neisser visual scanning task does not appear to have been used at all in
any dual task situations in the reported literature.

In a sleep deprivation study, Webb (1985) used an adaptation of the Neisser
task as part of a battery of tasks to determine cognitive performance in
sustained operations settings. Subjects searched an array of letters for
either an "X" or "Q" in contexts of either rounded letters (e.g., G, O, C,
D) or angular letters (e.g., V, N, K, Y). The test was not sensitive to
amount of sleep loss (with short naps allowed). The only significant
effect was found with sleep deprivation of older subjects (40 to 49 years).

Tuttle, Wood, and Grether (1976) used a battery of tests to measure per-
formance impairment of workers exposed to carbon disulfide (CS2). The
Neisser letter search task consisted of clusters of letters presented to
the subject on a sheet of paper for 20 seconds with instructions to iden-
tify and mark the predetermined letter(s) from the visual array. Two
trials each were given to search for single, dual, and four target let-
ters. The total number of target letters correctly identified during the
six trials was measured. Significant performance decrements were found in
the exposed group on the letter search task.

In a later study (Tuttle, Wood, Grether, Johnson, and Xintaras, 1977), the
same Neisser search task was used to determine the behavioral effects of
chronic perchloroethylene (PCE) exposures. No significant differences were
found in this experiment.

The same Neisser visual task was also used in a health survey of velsical
pesticide workers (Xintaras, Burg, Tanaka, Lee, Johnson, Cottrill, and
Bender, 1978). A set of cognitive tests were selected to evaluate the per-
formance of workers exposed to the pesticide leptohos. Unfortunately, the

lack of a comparison group (e.g., control/no exposure) made it impossible to clearly identify differences with any of these tests.

In summary, it appears that the Neisser visual scanning task may prove to be a sensitive measure of perceptual speed in drug testing if it is employed correctly (i.e., with proper experimental control).

TECHNICAL DESCRIPTION

Two versions of this test are available for use. The specifications for the two versions are as follows:

Lite Pen Version

At the beginning of each trial of the visual scanning task, a fixation point (character) is displayed on the top line of the screen three character positions to the left of center (one position to the left of where the array will appear). The purpose of the fixation character is to reduce the variability in the subsequent visual search time and to provide a preparatory time cue for the next stimulus presentation. The fixation character may be a right arrow, a dash, or an asterisk (roughly in that order of preference) depending upon character set availability and appearance.

The stimulus array consists of 25 rows and 5 columns randomly generated from the 25 letters "A" through "Z" excluding "K." The array is generated during the intertrial interval while the display is either blanked, displaying a feedback character, or a fixation character. One randomly selected character within the array is then replaced by the target letter "K," with the restriction that it may not occur within the first four rows of the last visible row. If the video adapter and monitor cannot handle a 25 line display then only the first 24 of 25 lines will actually be presented.

Once the array is displayed, it must not scroll, sweep, or be painted on the screen at a discernible rate but must appear within one frame interval triggered from the vertical sync pulse or equivalent. This implies that it

will reside in a different screen page than the fixation character, and
that if only one text page is available the fixation character may have to
be generated graphically. All letters are upper case, in white, on light
blue background, and dark blue border.

The instant the target is detected, the subject presses a button on the
button box, and then has 5 seconds to touch the target letter with a lite
pen. (Although the lite pen response might appear to be sufficient in and
of itself in this application, it contains inherent variability due to dif-
ferent physical movement times, different video beam scan times, and
usually to details of the "hit detection" circuitry or algorithm used.)

Keypad or Keyboard Version

The fixation stimulus and array presentation are the same as above. How-
ever, the occurrence of the button/keypress causes the array rows to be
immediately labeled with the numbers 01 through 25. These numbers are dis-
played to the right of the letter array after one intervening space.

The button is replaced by a designated key on the keypad or keyboard. As
with the button response, the subject has 10 seconds to respond. After the
response is made and the rows are numbered, the subject has 5 seconds to
enter the 2-digit target row number. A return or enter is not required,
and backspace correction is not allowed. In all other respects the second
digit entered serves the same function as did the lite pen responses in the
above.

Trial Specifications

Each trial begins with a 500 msec presentation of a visual fixation
point. When the fixation interval has elapsed, the stimulus array is dis-
played and the timer is started. The subject scans the stimulus array,
presses a button on the button box the instant the target letter is recog-
nized, and then has 5 seconds to touch the target letter with a lite pen
(or enter the proper row number on the keyboard).

Detection of a lite pen "hit" (or the second digit entered) initiates a
500 msec delay interval, optionally displays a feedback character during
this interval (in the same location used for the fixation character), then
blanks the array for 500 msec while displaying the next fixation charac-
ter. If a subject fails to detect the stimulus (no button response occurs)
within 10 seconds, or if no lite pen or keypad response is recorded within
5 seconds of the button response, the screen is blanked for 500 msec, and
the next fixation period begins. The task continues for 40 trials or
5 minutes, whichever occurs first. (The number of trials and test duration
may be varied by the experimenter.)

DATA SPECIFICATIONS

Each trial generates a stimulus code, a response code, and two time val-
ues. The stimulus code identifies the row and column of the target let-
ter. The response code identifies whether the response was a correct lite
pen response, an incorrect lite pen response, a late lite pen response, or
a late button response. These time values are replaced with the appro-
priate "deadline" values in the case of late (missing) responses. Summary
data requirements are: (1) task duration in seconds, (2) number of trials
completed, (3) number and percent correct ("late trials" count as errors),
(4) number of late button responses, (5) number of late lite pen responses,
(6) least square linear fit, derived from correct trial button reaction
times and target row locations, including: (a) slope of regression line
(scan time per row), (b) intercept (response time for "zero" rows), and (c)
squared correlation coefficient (r), (7) response times for correct
detections.

Raw summary data for keypad/keyboard version is analogous to the above.
The word "button" can be replaced by "detection key" and "lite pen" by
"second digit."

TRAINING REQUIREMENTS

The instructions to the subjects should be read to the subjects before the
start of the training trials. Following the instructions the subjects

should be presented with a minimum of 10 practice trials. The practice
trials will differ from the experimental trials in that following each
"hit" with the lite pen, a feedback character will be displayed indicating
the correctness of the response. The response or scanning time will also
be displayed.

During the practice trials the experimenter should carefully evaluate the
subjects performance in order to determine that the instructions are being
followed. For example, the instructions stress that the subjects scan the
array "quickly and accurately"; however, subjects may be sacrificing accu-
racy for the sake of speed, or they may be reaching the response deadline
too frequently. Furthermore, the experimenter should ensure that subjects
are scanning the array in the prescribed manner (i.e., from left to right
and top to bottom).

To summarize, the training phase for this test should consist of the
following steps:

1.   Read instructions to the subjects.

2.   Run practice trials and evaluate subjects' performance to ensure that
     the instructions are being followed.

3.   Repeat the practice trials if it appears that the subjects require
     additional practice with the test.

4.   Run the experimental trials. Note, if the tasks are being run over
     several sessions on this test, one may omit the practice trials after
     the first session.

INSTRUCTIONS TO SUBJECTS

This test examines your ability to make quick perceptual discriminations.
The computer will present you with a brief fixation character followed by a
25 row by 5 column display of letters from the alphabet. You are to scan
the array from left to right and top to bottom (in natural reading order)

for the presence of the letter "K." Scan the array as quickly as possible but be sure to identify the correct letter. Once you have detected the target letter (K), press the button on the button box, then picking up the lite pen, touch the "K" on the monitor with the pen. If no pen is available on this version of the test, after you press the button, the rows of the array will immediately be labeled with the numbers 01 through 25. Once the "K" has been identified, press the button, and then enter the two digit row number containing the "K."

Section 17

CODE SUBSTITUTION TASK (UTC-PAB TEST NO. 16)
(PERCEPTUAL SPEED, ASSOCIATIVE LEARNING ABILITY)

PURPOSE

This task is designed to tap information processing resources dedicated to the rapid encoding and associative evaluation of stimuli.

DESCRIPTION

The UTC-PAB code substitution task is derived from a paper and pencil version of the task contained within the Wechsler Adult Intelligence Scale (Wechsler, 1958), and is designed to assess associative learning ability and perceptual speed. A string of nine letters and a string of nine digits are arranged on a CRT display so that the digit string is immediately below the letter string. Each digit corresponds to a given letter. A test letter is then presented at the bottom of the screen, below the two coding strings. The subject is to indicate which digit corresponds to that test letter in the coding strings by pressing a designated key on a numbered keypad. The letter-digit associative pairings remain the same for the entire test.

BACKGROUND

The Code Substitution Test (also called the "Digit-Symbol" test) has been utilized as a psychological measurement tool for over 50 years (Pepper et al., 1985). The popularity of this test increased markedly upon its inclusion in the original Wechsler-Bellevue Intelligence Test (Wechsler, 1958) as a diagnostic test of intellectual speed. For the years that have followed, the Wechsler paper and pencil version of Code Substitution has been frequently utilized as an established metric of mental functioning due to the convincing data provided by Wechsler himself. High correlations were reported between the Code Substitution Test scores and overall IQ ($r = .67$ for ages 20 to 34; $r = .70$ for ages 35 to 49). Thus, the employment of

198

this task seems to represent a vehicle for assessing speed and efficiency of intellectual performance.

RELIABILITY

The reliability associated with all Wechsler-Bellevue subjects and scales was investigated by Derner, Aborne, and Castore (1950). Their subjects were classified as "normal adults" (n = 158). Once the task had been learned, simple test-retest reliability coefficients for the Digit-Symbol Test were all in excess of .70. This suggests that this test is of sufficient baseline reliability to potentially reflect performance decrements related to environmental stressors. However, research devoted to environmental effect on performance typically employ very extensive repeated measures designs. Thus, to be of value in performance assessment research, test-retest reliability must be established across several test sessions, as opposed to simple test-retest reliability.

Pepper et al. (1985) obtained performance data on the Code Substitution Task for 15 days from 19 Navy enlisted men, age 19 to 24. The subjects were given a 2-minute testing session each day. The performance metric utilized in these analyses was total items correct because subjects made very few errors and other measures were viewed as redundant (e.g., percent correct and reaction time would both be a reflection of total correct if performance is virtually errorless). The given scheme of letter/number correspondence was varied across days. Differential stability of performance was obtained by day eight. Cross-session reliabilities following this day were moderate and stable (r = .75). Thus, the authors concluded that the Code Substitution Test appears to be an "excellent candidate for assessment of environmental effects" based on these analyses of reliability and stability (Note: differential stability is characterized by high, stable test-retest correlations).

VALIDITY

Most discussions of validity that involve the Code Substitution Test stem from attempts to validate the complete set of Wechsler scales as a metric

of overall intelligence (Matarazzo, 1972; Wechsler, 1958). The validation tool utilized is the employment of correlational analyses among scores on several intelligence tests. Correlations among these various subtests and overall test scores tend to be high. It would seem, then, that these tests are presumably measuring the same cognitive abilities which can probably be combined into such a construct as intelligence. Of greater concern here, however, is the construct validity specific to the Code Substitution Test. While the correlations obtained by Matarazzo (1972) and Wechsler (1958) seem to validate the use of this task in the assessment of intelligence, the construct validity of each subtest, which carries more weight with respect to performance assessment, was not addressed in either publication.

Within the domain of performance assessment, the Code Substitution test is intended to specifically measure perceptual speed and associative learning ability. Validation of this test in terms of this construct was provided by Cohen (1957a, 1957b) who performed a series of factor analyses on the Wechsler-Bellevue subtests. Two principal factors were found to load on Code Substitution Test performance: A "perceptual organization" factor and a "memory" factor. The similarity between these factors and the test's construct is apparent. Thus, it can be stated with a sufficient degree of certainty that performance on this task taps into resources dedicated to perceptual speed and associative learning (or the retention of short term information). A potential link between the two parts of this construct, as pointed out by Cohen, might be the ability to filter out meaningless information at the perceptual level as well as the central (forming of associations) level of information processing. In summary, the test appears to assess the speed and accuracy with which an individual perceives new information and integrates it within the preestablished associative framework (Cohen, 1957a, 1957b). Also as discussed earlier, the ability to utilize these resources tends to stabilize for a given subject, permitting this test to be recommended for use as a diagnostic tool in performance assessment research.

SENSITIVITY

The sensitivity of the Code Substitution Task within the arena of perform-
ance assessment/stressor evaluation has not been widely investigated.
Because the test was originally developed as a subtest of the Wechsler-
Bellevue Intelligence Test, investigations of its sensitivity lie typically
within the clinical domain where the complete Wechsler battery is often
utilized. For example, Sax et al. (1983) utilized several Wechsler sub-
tests in an attempt to uncover cognitive predictors of the neurophysiologi-
cal correlates of Huntington's disease (HD). Declining performance on Code
Substitution was shown to be a function of the distance between the outer
tables of the skull and the caudate nuclei. This distance is typically
abnormal for HD patients, and it can be measured with a CT scan. Similar
findings are given considerable discussion by Wechsler himself (1958). He
cites several organic sources of decreased performance on all of the
Wechsler-Bellevue subtests. In general, any organic brain damage is seen
to impair performance to some degree on all subtests. Of special interest
here, however, is Wechsler's finding that "the greatest and most consistent
falling off (of performance) is on the Digit Symbol Test" (p. 174).
Wechsler also cites similar performance decrements for schizophrenia and
dementia praecox patients. Thus, organic brain damage is heavily reflected
in Code Substitution performance, indicating that this task is potentially
sensitive to any impairments which may beset an individual.

Also included in the clinically oriented research are sensitivity data
associated with sex and age. The effects of these variables could also be
brought to bear in the evaluation of environmental stressors and, thus, are
of considerable interest here. In general, females perform slightly better
on this task than males, and performance across all subjects tends to
decline steadily with age following 30 years of age (Wechsler, 1958). It
is important to bear these facts in mind when utilizing the code sub-
stitution task in any given area of research to avoid confounding these
factors with the potential sources of variation of interest and, thus,
ensure appropriate interpretation of obtained results.

As has been mentioned, the Code Substitution paradigm has not historically been employed in the study of environmentally introduced stressors. However, Pepper et al. (1985) introduced this task to such a paradigm to assess its potential as a diagnosic tool in the domain of performance assessment. As their subjects were six U.S. Coast Guardsmen, the variable involved was tolerance to sea motion. Motion-induced nausea was shown to produce a pattern of decrement in Code Substitiution performance similar to those associated with other perceptual/motor tasks (Wike, et al., 1979; Wiker and Pepper, 1978). This finding indicates that this task is potentially sensitive to the effects of environmental factors as well as organic factors. It seems, then, that the UTC-PAB version of the Code Substitution Task is sufficiently sensitive to be utilized as a diagnostic tool within the domain of performance assessment/environmental research.

TECHNICAL DESCRIPTION

The coding string remains displayed on the screen for the duration of a test session. Each test display consists of a string of nine randomly selected letters, and the digits 1 through 9 are strung directly underneath the letter string. Letters and digits are randomly paired for each test and order is randomly assigned in the coding string.

A single trial consists of the presentation of the probe letter to which a subject is to respond by pressing the key that corresponds to the appropriate digit. There are 30 trials per test session. There is an inter-stimulus interval (ISI) of 500 mesc between the subject's response and the presentation of the next probe letter.

The coding string is centered on the screen. The letters and digits are 2.0 cm in height and the letters are capitalized. The letter string is displayed 1.25 cm above the digit string and a given digit is located directly below its corresponding letter. The probe letter is designed to match the graphic features of the corresponding letter in the coding string. The probe is horizontally centered 6 cm below the bottom of the coding string. The probe remains on the screen until the subject makes a response.

202

If the subject makes a response during the ISI the screen will blank (i.e., the coding string will be removed), and the message "do not press a response key before the test letter appears" is displayed for 5 seconds. The coding string will then be redisplayed and the test will proceed normally. The response manipulandum is a numeric keypad which is separate from the keyboard. The subject responds by pressing the appropriate digit on the keypad.

DATA SPECIFICATIONS

The response time (recorded with less than 1 msecond error for each trial) and the actual correct response are recorded for each trial. Summary statistics are: (1) mean and median response times over the 30 trials, (2) range and variance of the response times, and (3) total number of correct responses. In addition, an option is available that allows examination of test performance on a trial by trial basis for each subject with each response time, correct response, and subject's response displayed.

TRAINING REQUIREMENTS

Trainig consists of the presentation of a coding string followed by 10 trials. The procedure is essentially the same as for the experimental trials, with the exception that subjects are provided with feedback during the training trials. If a subject responds inappropriately during training, the following message is displayed: "That was an incorrect response. The correct response was _____ (the correct code digit)." This message remains on the screen for 5 seconds. Then, the same probe letter is presented again. This procedure continues until all 10 trials have been correctly completed.

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

You will be presented with a row of letters across the screen. Directly below this is a row of numbers. The rows will be arranged so that the number directly below a letter is called the "code" for that letter. Your task is to learn the codes for each letter. A series of test letters will be presented, one at a time, at the bottom of the screen. These test letters are all taken from the letter row. Your job is to enter the digit on the keypad that is the "code" for that letter. For example, if the letter "J" was right above the digit "7," then "7" is the code for "J." When the letter "J" appears at the bottom of the screen, you should press "7" on the keypad. Try to respond as quickly upon the presentation of the test letter as possible without making any errors.

VISUAL PROBABILITY MONITORING TASK (UTC-PAB TEST NO. 17)
(SPATIAL SCANNING/SIGNAL DETECTION)

PURPOSE

The purpose of this task is to test perceptual resources devoted to scanning and detecting of visual signals.

DESCRIPTION

In this test, the subject is presented with a CRT display of dials and instructed to monitor the movement of a pointer located beneath each dial (Figure 13 shows a representation of the dials). Under normal conditions, the pointer moves from one position to another in a random fashion to simulate the pointer fluctuations on an actual dial. At unpredictable intervals, the pointer begins to move nonrandomly, staying predominantly to the left or right half of the dial. These biases in pointer movement are the targets or "signals" to which the subject is instructed to respond. The subject's job is to detect the presence of a "signal" and press the appropriate response key after which the biased dial will return to the original random pointer movement.

The test includes three task demand levels based on the number of dials that are displayed at any given time and the discriminability of the signals. A single test trial consists of 3 minutes of continuous monitoring and only one signal can be present at any given time. Signals may occur at any time within a trial with the restriction that a minimum of 25 seconds separates the offset of a signal and the onset of the next signal. Test trials typically contain two or three signals. In conditions where three or four dials are monitored (Moderate Task Level and High Task Level) the dial on which any signal will be displayed is randomly selected.
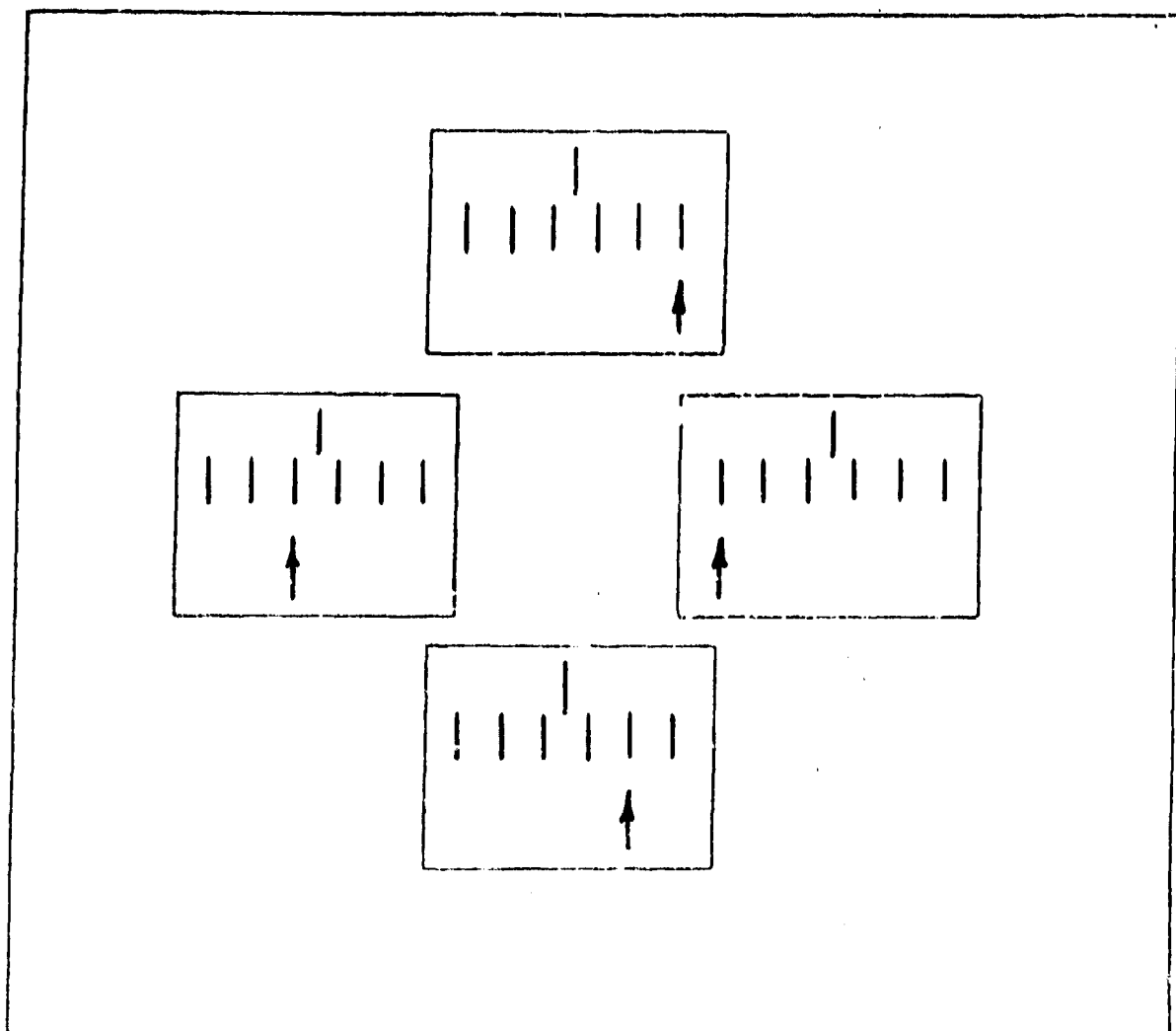
Figure 13. Probability Monitoring Display

206

BACKGROUND

The UTC-PAB version of the visual monitoring test was derived from a task developed by Chiles, Alluisi, and Adams (1968). The Chiles et al. (1968) task involved the monitoring of four meters for the presence of nonrandom fluctuation; however, unlike the UTC-PAB version, number of dials and signal discriminability was not varied. Furthermore, the dial monitoring task was performed concurrently with two other monitoring tasks (auditory vigilance and warning light detection).

Dial monitoring tasks have been used by other researchers (e.g., Carpenter and Conrad, 1953; Conrad, 1955); however, the procedure and display differed from the UTC-PAB version. For example, Conrad (1955) presented subjects with 4, 6, 8, 10, or 12 dials which consisted of a revolving pointer and marks at the 6 o'clock and 12 o'clock positions. The pointers on the dials stopped at a mark unless the subject pressed a button corresponding to the given dial. The subject's task was to keep all of the pointers moving. Conrad found that as the number of dials increased the number of stops per minute and average stopped time increased. Furthermore, recovery from errors (starting stopped dials) was more difficult as the number of dials increased. The subjective reports indicated that one was more "put off one's stride" by an error when the load was high (more dials) than when it was low.

Warm, Wait, and Loeb (1976) employed a task where subjects monitored a visual display for occasional increments in horizontal movements of a bar of light. This task is similar to the present test since subjects had to detect changes in the horizontal fluctuations of a vertical line segment (i.e., similar to a one dial condition). The results indicated that the detection probability was directly related to the amplitude of the increments in movement (2mm and 8mm changes for the low and high amplitude conditions, respectively) and inversely related to background events (the frequency of nonsignals occurring over time). Furthermore, the detection of signals at the low amplitude was enhanced by restraining subject's head.

207

The above research indicates that performance in a Visual Probability monitoring Task is directly related to signal amplitude and inversely related to the number of neutral events that a subject must monitor in search of critical signals. Furthermore, the number of signal sources (Conrad, 1955) is also inversely related to monitoring performance.

The UTC-PAB version of the Visual Probability Monitoring Task was designed with the following guidelines: (a) The Visual Probability Monitoring Task is based on a model of human information processing which posits three primary stages of processing and associated resources dedicated to perceptual input, central processing, and motor output or response activities (Shingledecker, 1984). The above model is based on multiple resource (Wickens, 1984) and processing stage (Sternberg, 1969) theories of human information processing. The Visual Probability Monitoring Task is assumed to tap visual perceptual resources and at the same time engage minimal central processing and output resources. (b) The actual nature of the present task was determined empirically. The number of display sources (one, three, or four dials) and stimulus discriminability (95/5, 85/15, and 75/25 percent probability bias) were factorially combined during the task development phase. These two variables were manipulated since they logically affect visual information processing (e.g., affect the signal to noise ratio). The three levels of task demand represented in the present version of the task were those combinations of number of signal sources and stimulus discriminability that were statistically different from each other and represented increasing level of task difficulty (e.g., longer response latencies and increases in error rates). The results from the test development phase are presented in Figure 14 (Shingledecker, 1984). Obviously the above procedure confounds the factors of numbers of signal sources and stimulus discriminability; however, the goal of the task developers was not to model the effect of the above variables on performance but to develop a task which posed reliably different demands on the systems (human) ability to process visual input.

In summary, the UTC-PAB version of the probability monitoring test appears to tap resources principally related to visual perceptual processing. Also, the fact that this test presents three increasingly difficult levels
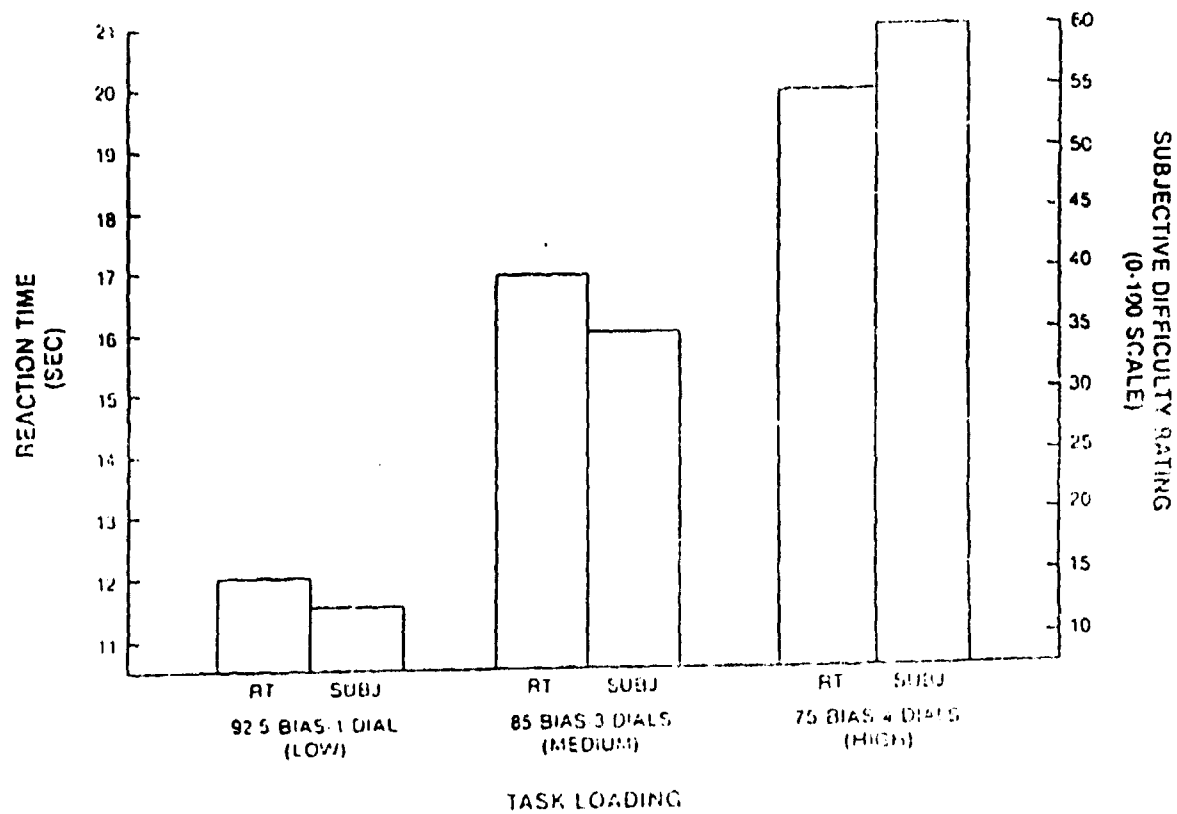
Figure 14. Mean Reaction Times and Subjective Difficulty
Ratings for Probability Monitoring Conditions

of task demand makes it amenable for drug research. For example, research on the effects of CO on cognitive performance (e.g., Johnson et al., 1974; Putz, 1979) showed detrimental effects for this drug when subjects were performing the high demand (or difficult) condition but not the low demand condition (this was specially true for dual task procedures). In addition, this test may be readily incorporated into a dual task procedure.

RELIABILITY

Research by Shingledecker (1984) with this task indicates that there is very little practice effect; that is, subject's performance is relatively stable at the start of testing. However, three to four 3-minute practice sessions are recommended in order to assure steady state performance.

Additional reliability data are available from Chiles et al. (1968). In this study two experiments were conducted to examine the test retest reliability (24 hours) of a meter monitoring task. In the first study (N = 15), reliability coefficients of .78 and .81 were determined for percent correct detections and reaction time to correct detections, respectively. The second study found reliability coefficients of .97 for percent correct detections, and .95 for reaction time to correct detections (N = 25). A study by Chiles, Bruni and Lewis (1969) examined the test-retest reliability of the visual probability monitoring tasks under three different signal rates; (a) training rate of 15.5 signals/hour, (b) slow rate of 9.4 signals/hour, and (c) fast rate of 20.6 signals/hour. The correlation coefficients for these conditions are presented on Table 11. (Note: N = 10 for this study.) However, a study by Chiles, Jennings, and Alluisi (1978) reported a reliability coefficient of .59 for reaction times in the meter monitoring task.

The above reliability data indicate that the visual probability monitoring task yields reliable response measures over time. This was especially true for the fast presentation rate in Chiles et al. (1969). However, these data may not apply directly to the UTC-PAB version of the task since it differs procedurally from the Chiles et al. (1968) version. For example, the UTC-PAB version varies the number of signal sources (one, three, or

210

TABLE 11. CHILES, BRUNI, AND LEWIS (1969) VISUAL PROBABILITY
MONITORING RELIABILITY DATA FOR RESPONSE TIME MEASURES

		Rate	
	Training	Slow	Fast
Spearman r	.77	.74	.92
Product-moment r	.88	.53	.96

four dials) and the discriminability of the signals, whereas these
variables were held constant by Chiles et al. (1969).

VALIDITY

As stated earlier, this task was based on a model of multiple resources
(e.g., Wickens, 1984). However, relatively little dual task research has
been conducted with the UTC-PAB version of the test. Shingledecker, Acton,
and Crabtree (1983) examined performance on this monitoring task when it
was time shared with the Michon tapping task (see Manual No. 19). The
Michon tapping task did not interfere with performance on the monitoring
task. The Michon tapping task is assumed to principally tap resources
associated with response timing and, therefore, should not interfere with a
task that does not place heavy demands on this resource. This negative
finding supports the notion that visual probability monitoring is a
resource specific task (e.g., visual processing resources); however, dual
task research which demonstrates performance decrements in visual monitor-
ing is needed (e.g., research that combines the visual monitoring task with
another task that purports to measure visual processing).

Chiles (1977) examined performance in the visual monitoring task when it
was combined with other tasks (Table 12). As can be seen in Table 12, the
meter monitoring task (e.g., visual probability monitoring) was always com-
bined with an additional monitoring task (e.g., warning lights). In
addition, this pair of tasks was always combined with two other additional
tasks. Chiles (1977) found that responses on the meter monitoring task
were fastest during interval one, next fastest in intervals two and four,

211

and slowest during the third interval. The difference in detection time between intervals two and four combined versus interval three was about 10 seconds. However, the difference in detection times for intervals two, three, and four combined versus interval one was about 60 seconds.

TABLE 12. TASK COMBINATIONS FOR A ONE HOUR TASK SCHEDULE FROM CHILES (1977)

15-Minute Intervals	1	2	3	4
Warning Lights	X	X	X	X
Meter Monitoring	X	X	X	X
Mental Arithmetic	X	X		
Tracking, Two Dimensional	X			X
Problem Solving		X	X	
Pattern Discrimination			X	X

(NOTE: An "X" indicates that the task was present.)

The above pattern of results is difficult to interpret with respect to the effect of additional task on meter monitoring performance. However, it appears that performance on the meter monitoring task will be disrupted when heavy demands are placed on working memory processing (e.g., mental arithmetic and problem solving) or when an additional visual processing task is added (pattern discrimination). Performance on the monitoring task was least affected when mental arithmetic and tracking were performed concurrently. A tracking task will most likely place heavy demands on motor output processing (similar to the Michon tapping task) and, thus, will not interfere with meter monitoring. Hall, Passesy, and Meighan (1965) found the same basic results when an Auditory Vigilance monitoring task was added.

The results of the above studies provide support for the idea that the UTC-PAB visual probability monitoring task taps resources associated with visual information processing. However, only one study used the present version of the task (Shingledecker et al., 1983) and the other studies always combined meter monitoring with additional monitoring tasks.

Dual task research that combines visual probability monitoring with such tasks as visual pattern comparison (visual information processing), mental arithmetic (working memory), and tracking or tapping (response output) may help to bolster this test's construct validity (e.g., a task that principally taps perceptual resources associated with the detection of visual signals).

SENSITIVITY

Research by Chiles and Jennings (1970) showed that performance on a meter monitoring task was degraded by the consumption of alcohol. However, the meter monitoring task was always combined with two additional monitoring tasks (light monitoring and choice reaction time to visual stimuli). In addition, Chiles et al. (1968) showed decrements in performance on the meter monitoring task as a result of sleep loss. Again, this experiment combined meter monitoring with other monitoring tasks.

Performance on the meter monitoring task appears to be sensitive to such factors as sleep loss and alcohol ingestion. However, it is difficult to predict to what degree the UTC-PAB version of the test will show sensitivity to environmental stress or drug status. The present version of the task has not been widely employed as a stand alone task or in dual task research.

TECHNICAL DESCRIPTION

A single test trial consists of 3 minutes of continuous monitoring. Test trials are equally likely to contain two or three signals. Signals may occur at any time within a trial with the restriction that a minimum of 25 seconds separates the offset of a signal and the onset of the next signal. In conditions where three or four dials are monitored, the dial on which any signal will be displayed is randomly selected.

When no signal is present, the pointer moves to each position with equal probability (1/6). When more than one dial is to be monitored, the pointer movement on each dial is independent of the others. Pointer position is

updated at the rate of two moves/second. Dials always appear in the same screen location (i.e., dial No. 1 is always located at the upper-center of the screen, dial No. 2 at the middle-left, etc.). In the single dial condition, dial No. 1 is displayed; in the three dial condition, dials one, two, and three are shown; and in the four dial condition, all four dials are displayed.

If undetected, a signal lasts 30 seconds and occurs over 60 pointer moves. When a signal occurs in the high discriminability condition, 57 of the 60 pointer moves appear on one side of the dial (95/5 percent probability bias); in the moderate discriminability condition, 51 of the 60 moves appear on the favored half (85/15 percent probability bias); and in the low discriminability condition, 45 of the 60 moves occur in the bias direction (75/25 percent probability bias). Within these constraints, however, pointer movement is randomly determined. Biases are equally likely to appear on either half of the displays and on any given display.

Three significantly different task demand levels are produced by the following task conditions: (a) low demand--one dial at the 95/5 percent bias level; (b) medium demand--three dials at the 85/15 percent bias level; and (c) high demand--four dials at the (75/25) percent bias level.

Trial Specifications

This test does not present discrete stimuli for responses, rather signals are presented for 30 seconds or until a response is recorded by the subject. Each 3-minute trial will contain two to three signals and the sequence of events for each signal period is as follows: (a) a signal bias is produced on one of the dials (only one dial will be biased at any given time), (b) the subject presses a key which corresponds to the location of the biased dial, (c) if the key pressed corresponds to the actual location of the biased dial or if the dial has been biased for 30 seconds, the biased dial will go back to its "normal" rate of fluctuation. The above sequence of events is repeated for each signal period.

DATA SPECIFICATIONS

For each 3-minute trial the following information will be recorded: (a) signal condition (low, medium, or high demand condition), (b) trial start (time 0), (c) onset time for each signal, and (d) responses entered by the subject (e.g., dial number), along with the elapsed time of response occurrence in msec from the start of the trial.

The following summary statistics will be calculated for each 3-minute trial: (a) number of signals presented, (b) number of correct signal detections, (c) number of missed signals, (d) number of false alarms, and (e) reaction time for each correct signal detection.

Note: the present rate of signal presentation per 3-minute trial is rather low (e.g., two to three signals). This will result in a small number of responses which will make the use of parametric statistical procedures questionable. Increasing the rate of signal presentation per 3-minute trial may remedy this situation (research has been conducted at AAMRL on a visual monitoring task with a faster rate of signal presentation; however, the results of this research have not been published). Research by Chiles et al., 1969) has shown that increasing the rate of signal presentation in a meter monitoring task increases the reliability of the response meas-ures. Perhaps the above suggestion will lead to the development of a visual monitoring task which yields behavioral measures that are reliable and parametrically sound.

TRAINING REQUIREMENTS

Subjects should be initially introduced to this test by presenting them with the instructions. Following the instructions, the subjects should be presented with a minimum of two to three practice sessions per demand condition. In addition, during the practice trials the subjects should be cued as to the presence of a dial bias. The detection of a dial bias, specially at the high demand level, will require subjects to become familiar with the appearance of a dial bias before testing can proceed.

During the practice trials the experimenter should stress the fact that the subjects not respond until they are certain that a signal is present. In other words, the strategy of responding more frequently than necessary to avoid missing signals is undesirable.

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

In this task you will be monitoring a number of displays which are intended to have the appearance of electromechanical dials like those on a machine. The dials consist of six pointer positions and a pointer which appears below the positions and moves from one to another. Under normal conditions the pattern of pointer movement is random. The pointer is equally likely to move to any position. Periodically the pointer movement on one of the dials will become nonrandom, such that the pointer will tend to stay on one side of the dial more than the other. Your task is to watch the dials carefully for nonrandom or "biased" patterns of pointer movement. Biases in pointer movement are called "signals." If you think you see a signal, press the button on the keypad that corresponds to the dial. When you correctly respond to a signal, it is eliminated and the pointer goes back to moving randomly again.

Monitoring periods last 3 minutes each. You start the monitoring period
when you are ready by pressing any of the response keys. During each
3-minute period, you can expect to see two or three signals (biases). If
you don't respond, a signal lasts for 30 seconds, so there is ample time to
make a decision before responding. When you make a response, the computer
generates a tone to let you know that it was received. More than one sig-
nal may appear on a given dial during the 3-minute test period, but two
signals will never appear on different dials at the same time. Try to
avoid responding unless you are confident that a signal is present.
Responses to nonexistent signals are scored against you. The screen will
automatically go blank at the end of the monitoring period.

Two aspects of the monitoring task will vary from trial to trial. The
first is the number of dials to be monitored. You will be monitoring
either one, three, or four dials at a time. The other variable is the pro-
portion of time the pointer spends on the favored side of the dial when a
signal occurs. In the one dial condition, the pointer will stay under the
favored half of the dial 95 percent of the time, and will appear on the
nonbiased side only 5 percent. In the three dial condition, this propor-
tion is more equal: 85 percent of pointer moves will be on one side, and
15 percent on the other. The proportion of moves is most equal in the four
dial condition, 75 to 25 percent. The effect of equalizing the proportion
of time spend on each side of the dial occurs when no signal is present.
Therefore, a 75/25 signal tends to look more like random, nonsignal pointer
movement than an 85/15 or 95/5 signal.

Section 19
TIME WALL (UTC-PAB TEST NO. 18)
(TIME ESTIMATION)

PURPOSE

The purpose of the time wall task is to test a subject's ability to esti-
mate the time at which a target, moving at a constant rate, will have trav-
eled a predetermined distance. That is, on each trial the subject must
integrate the available speed and distance information in order to cor-
rectly anticipate the time at which the target reaches a certain spot on
the screen.

DESCRIPTION

The UTC-PAB time wall task is a nonverbal time estimation task in which a
small object moving at constant velocity passes behind an opaque barrier
and the subject must estimate the moment when the object will reappear.
The time wall differs from a number of other time estimation tasks in that
discrete mediating responses such as counting or tapping are of no direct
obvious aid. In this implementation, movement is vertical rather than hor-
izontal for purposes of visual field symmetry. The barrier contains a hole
or notch the same shape and size as the object, and the subject estimates
the moment when the entire notch will be filled. This implementation uses
a nominal 10-second time interval.

BACKGROUND

The time wall task originated in a group of experiments conducted at the
Armstrong Aerospace Medical Research Laboratory in order to determine the
effects of noise on vigilance and time judgements (Jerison, Crannel, and
Pownall, 1957; Jerison and Arginteanu, 1958). The first experiment dem-
onstrated an effect of noise in a rate projection situation in which sub-
jects judged the time required for a target moving at a constant speed to
traverse a part of its route in which it was invisible. It was shown in
the experiment that a noise program in which it was quiet during the

visible portion of the target's course and noisy (110 db SPL) during the invisible portion when the subjects made their judgements, resulted in judgement times displaced upward (overestimating) relative to judgement times under other noise programs including the reverse (noise for visible, quiet for invisible) program.

In the Jerison and Arginteanu (1958) study, the same rate projection task was used; however, five different speeds and four noise levels were factorially combined. The four noise programs were noise throughout (108.5 db) (NN); quiet throughout (QQ); noise when the target was visible followed by quiet when the target disappeared (NQ); and quiet when the target was visible followed by noise when the target disappeared (QN). The five target rates were: 0.8, 0.4, 0.2, 0.1, and 0.05 inches per second.

A small target pip was generated and displayed on a 21 inch television tube, and movement was always across a left to right horizontal path (Figure 15). The target could be seen for four inches across the path and was invisible for two and a half inches thereafter. The subject responded by squeezing a trigger on an ordinary aircraft control stick at the time the target was estimated to be at a marked location. All subjects received all of the combinations of rate and noise programs in random orders.

In calculating the results, the judged time interval was divided by the correct time interval. This new measure was used to allow the comparison of subjects performance for each of several "correct" intervals. The shortest "correct" interval (3.12 seconds) was obtained when the fastest rate (.8 inches per second) was used. The other correct intervals were: 6.25, 12.25, 24.42, and 48.12 seconds for .4, .2, .1, and .05 inches per second respectively. The effects of rate and of noise were both found to be highly significant, the interaction was not. When the judged interval/ correct interval is plotted against the correct interval for each noise program, the resulting curves show that all time intervals were overestimated (Figure 16). None of the four curves cross under the "indifference interval" (the point at which the judged interval equals the correct interval). The typical result of a time judgement experiment is often summarized as indicating that short time intervals are overestimated
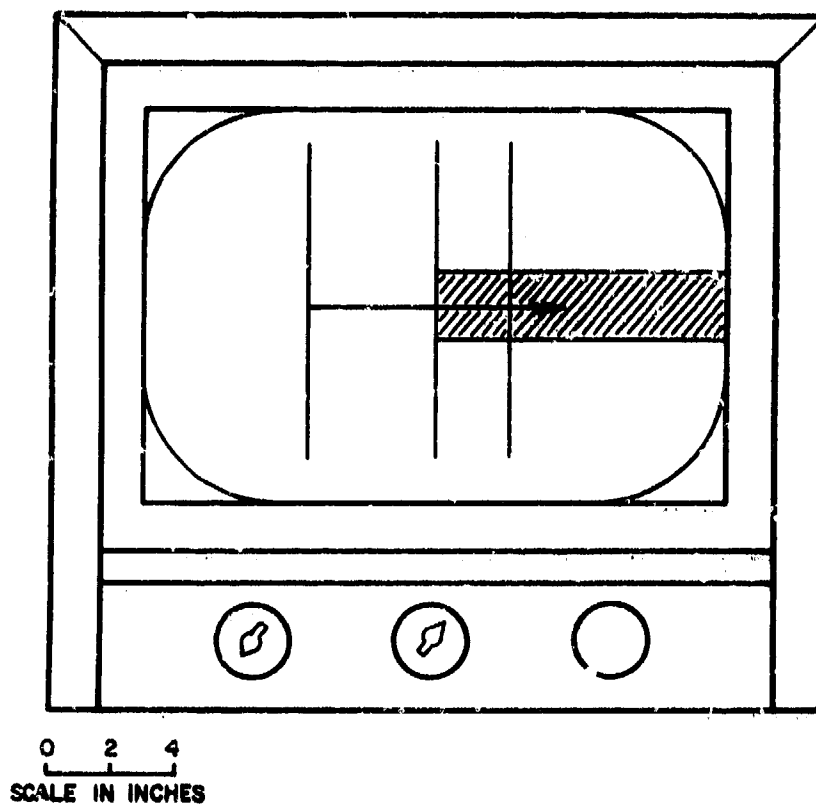
Figure 15. The Stimulus Display [The Arrow Represents the Path of the Moving Target Pip. The Portion Under the Shaded Area was Invisible. (Jerison and Arginteanu, 1958)]

220
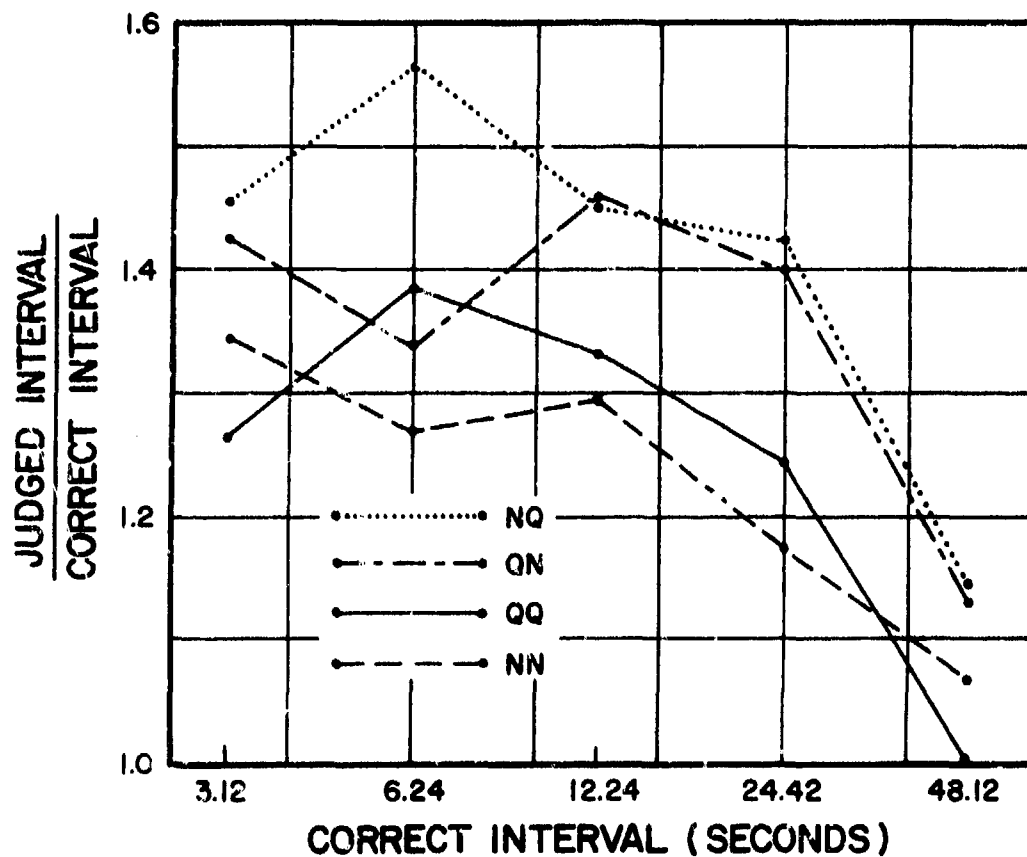
Figure 16. Relative Error for Various Rates of Movement of the Target Pip [Rates are Converted into Correct Interval Measures which Reflect the Duration of the Invisible Portion of the Target's Course from the Disappearance Point to the Vertical Cross-Hair. Correct Responses would Yield a Value of 1.0 on the Ordinate. (From Jerison and Arginteanu, 1958)]

221

and long time intervals underestimated. The downward sloping curves of the results indicate that all of the intervals were overestimated, though the amount of overestimation became less for longer intervals.

The four noise programs significantly differentiate performance. The QN and NQ programs produced a greater degree of overestimation than did the NN and QQ programs. That is, noise had an effect in terms of whether it was steady or whether its level changed at the time of disappearance of the target.

The time wall task is different from other time estimation tasks in that the passing of time is anticipated based upon other information such as rate and distance that is available to the subject. This is a relative judgement since the subject has witnessed the amount of time the target had taken to travel the visible distance. There is no task interference during the judgement interval. In a typical time estimation experiment, the subject's task is to estimate how much time has elapsed while performing another task. In this case, time is judged on more of an absolute basis, without other helpful information. These researchers are interested in how different levels of workload imposed on the operator affects his perception of the passing of time. Another major difference between the two paradigms is in how the time interval is determined. In most time estimation experiments, the length of time of the interval to be estimated is selected by the experimenter and the subject attempts to determine what the interval was. In the time wall paradigm, the subjects themselves determine the length of the time interval based upon the stimulus condition.

An experiment by Aitken and Gedye (1968) provides a good example or a typical time estimation experiment and its results. In the experiment, eight Air Force pilots were isolated for four intervals of 10 minutes. During two of the intervals the pilots were required to perform a simple tracking task, while in the other two they were not required to do anything. In addition, on one occasion for each task condition they were exposed to distracting stimulation (noise). The subjects were to estimate the duration of each interval and indicate how alert they had been during it. The results obtained were typical of time estimation observations in general.

The apparent duration of the interval was increased by the presence of the distraction and decreased by the performance of the concurrent task.

Hicks, Miller, and Kinsbourne (1976) critically reviewed procedural differences in time estimation experiments. These authors distinguish between the information presented "to" subjects during an interval and the information processed "by" subjects during an interval. When processing of the stimulation is not required of the subject, judged time is usually an increasing function of the number of stimuli or the complexity of the stimuli that occur during an interval. The function changes, however, when the processing of information is required. When the subject must process the stimulation presented or perform some concurrent task during the interval, the judged time then decreases with the activity or information processing.

To summarize, the time wall task agrees with other time estimation tasks in that shorter intervals tend to be overestimated and the presence of a distraction (e.g., noise) tends to increase the assessed duration. The time wall, however, possesses several distinctions from other time estimation paradigms. Time estimation for this task is performed "on line" or during the actual occurrence of the interval. This is opposed to the more common technique of making an estimation after the interval has elapsed. Time wall judgements are relative estimates of time. Subjects can use the rate and distance information from the visible portion of the trial as an aid or predictor of the invisible portion. No time reference is usually provided in other paradigms. Finally, by pulling the trigger, the subject is terminating the interval for that trial. Although the "correct time interval" may have been surpassed, the subject terminates the trial. In other paradigms, the interval is terminated by the experimenter at a predetermined time. Because of these differences, the time wall task has been classified as a test of rate projection or time anticipation, and not strictly time estimation.

RELIABILITY

When experimentally testing for the effects of environmental factors, measurements over several days and times are usually required. Therefore, in

environmental research, it is important that the measure consistently
demonstrates the same outcome over these several applications. This test-
retest reliability has not yet been determined for the time wall task.
However, Jerison and Arginteanu (1958) did examine trends resulting from
repeated measurements on the same subject over three days. Successive
blocks of 20 trials (two each day) were used in the analysis to display
trends due to repetition of the task. The results indicate an unmistakable
upward trend in judgement times over blocks. This trend continues across
days of work. Both blocks and days were significant in effecting time
estimation. These results indicate that performance on this task does not
stabilize readily. In fact, subject's performance worsens over time as
they overestimate more each day. Thus, many practice trials might be
required on this task for performance to stabilize. The use of feedback
might also alleviate much of the tendency to overestimate and lead to
higher reliability at much lower levels of practice. In summary, more
research is needed to adequately determine the reliability of the time wall
test and at what point performance stabilizes.

VALIDITY

In typical time estimation studies, a person is required to judge the
length of time that has elapsed over a period in which some activity may or
may not have been performed concurrently. The judged interval may range
from 40 seconds up to and beyond 10 minutes. These studies have generally
shown that when no processing is requ   d, increases in stimulus complexity
produce monotonic increases in judged time. However, when processing of
information is required, judged time decreases with activity or information
processing.

In the time wall task, the person does not judge the length of time of an
elapsed interval, but more correctly attempts to project the rate or speed
at which the target is traveling. From this rate projection, the person
must anticipate the short interval of time the target needs to travel a
known distance. Thus, the time wall task is qualitatively different from
other time estimation tasks and requires different resources than those
used to judge absolute time intervals. The time wall utilizes resources

associated with the integration of rate of motion and distance information, not necessarily the passing of time. There is evidence that the verbal estimation of short intervals (i.e., $\leq$ 10 seconds) involves partially different processes than the verbal estimation of longer intervals (Hicks and Miller, 1976). Although it is clear that the time wall task requires other resources, in addition to those used in judging absolute intervals of time, precisely what other resources are required in the task are speculative until more research involving the task is conducted.

SENSITIVITY

The sensitivity of a test is determined by how well a given manipulation reflects a change in performance. For the purpose of this battery, it is important that the test shows sensitivity to drug effects. The time wall task has been used in one study to determine drug effects. This study will now be described.

Seppala and Visakorpi (1983) investigated the effects of oral atropine on a variety of psychological and physiological tests. These measurements were made before a single oral dose of atropine (.85 or 1.7 mg, or a placebo), and 1, 2, and 4 hours after it. Measures taken included flicker recognition, reaction time, short term memory, coordination, time anticipation, and standing steadiness. The version of the time wall task used in this study was named the Time Anticipation Reaction Test (TART). In the TART, the test persons had to estimate the time in which a small round light, gliding at a speed of 16.8 cm/seconds (6.6 inches per second), would need to pass a certain wall. The test persons indicated their estimation by pressing a key. The measure obtained was the coefficient of variation (CV) where $CV = \frac{SD}{Mean} \times 100$. The CV was calculated from the mean and standard deviation (SD) of trials after two training trials. Ten successive estimations were computed. The target traveled behind a wall for a distance of 13.75 inches. The correct interval to be anticipated was 2.08 seconds.

According to the analysis, atropine tended to have no effect on time anticipation. However, atropine distorted the distribution of the time anticipation scores so that the lower dose produced a somewhat flattened

distribution and the higher dose a more flattened even distribution. An insight to the test persons' individual responses revealed that the distributions were distorted because the initially "fast estimators" (mean anticipation times: 1.55 to 1.75 seconds) reacted still faster and the initially "slow estimators" (mean anticipation times: 2.34 to 2.56 seconds) reacted even more slowly after the drug. The test persons, whose anticipation times (means: 1.99 to 2.18 seconds) were initially near to the correct anticipation time (2.08 seconds), were not affected by either dose of atropine.

Although time anticipation was not found to be sensitive to atropine at the rate of 6.6 inches per second, other rates may reveal different functions. Jerison and Arginteanu (1958) used much slower rates of under 1 inch per second. Since the distance to be traveled was short, these rates produced time intervals between 3 to 12 seconds. Perhaps different time intervals to be judged are differentially sensitive to environmental factors. Therefore, a sensitivity study employing a number of rates may provide a better assessment of the sensitivity of this task.

TECHNICAL DESCRIPTION

The barrier (wall) occupies the lower third of the display area. The notch (missing brick) is centered along the wall's bottom edge. The moving object (falling brick) emerges from the top of the display area and descends at a constant velocity such that its leading edge would reach the bottom line of the display at a precisely known time (nominally 10 seconds). The brick appears to pass behind (or into) the wall, after which the timer continues to run but nothing else occurs until the subject responds or a deadline elapses.

Target distance shall be determined by the VDT screen dimensions. Rate of the target depends on time and distance values. However, several rates resulting in judgement intervals of between 2 to 10 seconds would be preferable. The brick and notch are identical small squares whose size may have to be determined after initial viewing on the selected monitor and video adapter. Tentative dimensions are three-sixteenth inch squares.

Monitor colors to be used in the task are a dark blue border, light blue sky (upper two-thirds of display), a light blue notch, and white wall. If that large an expanse of white appears adversively bright, as is often the case on monochrome displays, then a wall color should be selected from the available palette to provide good color contrast but with a subjective brightness approximately equal to the light blue.

Trial Specifications

Each trial in the task begins when the brick emerges from the top of the screen and descends at a constant velocity behind the wall. The subject estimates the brick's transit time (the time at which the target should fill the notch at the bottom of the wall) and presses any button on the button box. Feedback that an acceptable response has been made is provided by instantly filling the notch with the wall color. After 500 mseconds, the notch reverts to light blue and a new brick begins to emerge from the top of the screen.

If a button is pressed before the brick has passed completely beyond the upper edge of the wall, then the trial continues without visible change but an "extra" response is counted. If no acceptable response occurs within 30 seconds, then a beep is sounded and the next trial begins 1 second later. The task continues for 10 trials or 300 seconds, whichever occurs first.

DATA SPECIFICATIONS

Each trial generates at least one time value and a response code indicating whether the response was acceptable, an "extra," or was timed out by the deadline. Times are measured from the start of each trial and will usually have values around 10 seconds. Recorded values for deadline occurrences are set equal to the deadline value itself (i.e., 30 seconds). These times may be recorded as their absolute values or assigned differences from the calibrated (nominally 10 seconds) standard.

The following summary statistics are computed and stored: (1) calibrated standard time value (not necessarily 10 seconds), (2) total elapsed time (task duration in seconds), (3) number of trials completed, (4) number of "extras," (5) number of time outs (deadlines), (6) constant error (mean estimate minus standard), (7) proportional error (mean estimate as a percent of standard), (8) variable error (standard deviation of the estimates in mseconds), and (9) coefficient of variation (standard deviation/mean estimate x 100).

TRAINING REQUIREMENTS

The instructions should be read to the subjects before the start of the training trials. No training requirements have been established. However, in the experiment by Jerison and Arginteanu (1958), time estimations were not stabilizing after the third day, or after six blocks of trials. Therefore, at least six practice blocks should be performed. If only a single rate is used, performance may stabilize earlier. The experimenter should monitor the subject's performance to determine at what point time estimation values are stabilizing.

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

This is an experiment to see how well you can estimate the speed of a moving square target. The target will always start at the top of the screen and descend at a constant rate toward the bottom. After the target is two-thirds of the way down, it will pass behind a wall and become invisible. Your task is to press a button at the exact moment the moving target would pass through the notch marked at the very bottom of the display. In making this judgement, you are not to count or use any other rhythm method to facilitate your judgement. Instead, follow the target with your eyes and imagine it continuing straight down behind the wall to the notch. After you have pressed the button, you will receive feedback as to where the target actually was and whether you over or underestimated the time interval. A half second later, the next target shall emerge from the top. The task continues for 10 trials or 5 minutes, whichever occurs first.

INTERVAL PRODUCTION TASK (UTC-PAB TEST NO. 19)
(RESPONSE TIMING)

PURPOSE

This task was designed to be used as a secondary task to measure demands
placed on motor output by a primary task (Michon, 1966). However, it may
be used as a stand alone test to examine the degree to which variables such
as drugs, environmental stress, and toxic substances disrupt manual
response timing.

DESCRIPTION

This test requires the subject to generate a series of time intervals by
tapping a finger key at a rate of one to three responses per second. The
subject taps with the forefinger of the preferred hand using a paddle
shaped key (approximately one and one-half inches by three inches). The
task is run in 3-minute trials and the subject is encouraged to maintain
equal time intervals by tapping at as regular a rate as possible. Inter-
vals are timed from the onset of one response to the onset of the next
response and intervals of less than 10 msec are rejected as spurious input.

BACKGROUND

Michon (1966) developed the tapping task as an all purpose secondary
task. The secondary task method assumes that humans have a restricted
capacity for handling information. If this capacity is not fully engaged
by the particular task under concern, it should be possible to perform
some other task simultaneously. This conceptualization of processing
capacity assumes an undifferentiated pool of cognitive resources; however,
current theories of human information processing (Wickens, 1981) propose
that cognitive resources may be differentiated along such dimensions as
input (auditory, visual) and output (verbal, motor) modalities. This issue
will be discussed in further detail when reviewing the results of experi-
ments that have utilized the tapping task. At any rate, Michon proposed

that the major difficulty in performing two tasks simultaneously is essen-
tially a matter of temporal structuring of perceptual motor behavior.
Therefore, performance on a secondary task such as tapping (which requires
the timing of a motor response) can serve as an index of the processing
capacity not being utilized by the primary task.

The procedure for using the tapping task in a dual task experiment entails
two basic steps: (a) the basic tapping level (BTL) is determined for each
subject where tapping is performed alone, and (b) the loaded tapping level
(LTL) is determined where subjects are performing the tapping task in con-
junction with a primary task. The above tapping levels (BTL and LTL) are
measures of tapping variability. Michon (1966) recommended the following
formula for computing tapping variability:

$$IPT_{variability} = \frac{N}{T} \sum_{i=1}^{N} |\Delta t_i|$$

where N is the total number of intervals produced, T is the total time over
which data is collected, and Δt_i is the difference between successive
intervals. Lower values for the above formula indicate more temporally
regular tapping. In addition, the above measure of tapping variability is
superior to such measures as the standard deviation of interval duration
because it corrects for the partial dependence of error magnitude on
interval duration (Figure 17 shows sample computations).

Michon (1966) evaluated the effect of primary task performance on tapping
performance by computing what he referred to as Perceptual Motor Load
(PML). PML is computed with the following formula, PML = (LTL- BTL)/BTL.
As can be seen, a value of zero for PML would indicate that tapping was
performed at the same level under single and dual task conditions.

Michon (1966) proposed the tapping task as an inobtrusive, easy to learn,
stable, and sensitive secondary task. In addition, the proposed PML mea-
sure could serve as a metric for comparing a diverse set of primary
tasks. However, the tapping task has received relatively little attention
in the dual task literature. Table 13 presents a summary of dual task
research with the Michon tapping task.

TIME (sec)

| 0 | 10 | 20 | 30 | 40 | 50 |

A. || | || ||| | | | | | || |

$\bar{t}$ = 1.71   T = 25 $\Sigma |\Delta t|$ = 10
s = .70   N = 14

$$IPT = \frac{N}{T} \Sigma |\Delta t| = \frac{14}{25} (10) = 5.60$$

B. | | | | | | | | | | | | | | |

$\bar{t}$ = 3.43   T = 50 $\Sigma |\Delta t|$ = 20
s = 1.40   N = 14

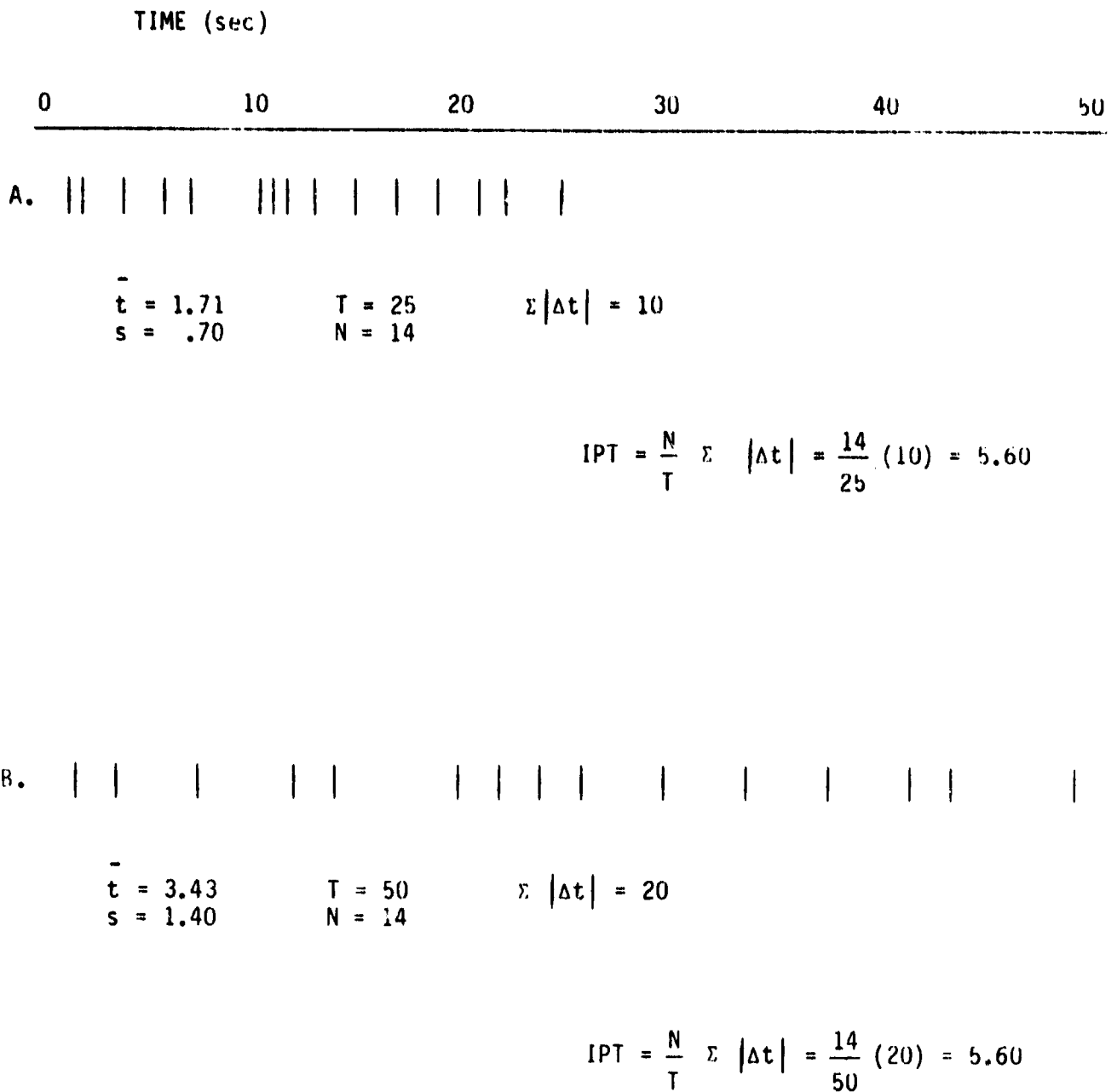$$IPT = \frac{N}{T} \Sigma |\Delta t| = \frac{14}{50} (20) = 5.60$$

Figure 17. Two Hypothetical Tapping Records [Record A is for a Series of 15 Taps Over a 25-Second Interval and Record B is for 15 Taps Over a 50-Second Interval. The Vertical Lines Under the Time Line Represent Taps. Note: S is the Standard Deviation of the Tapping Intervals and IPT is the Measure of Tapping Variability Recommended by Michon (1966).

TABLE 13. SUMMARY OF DUAL TASK RESEARCH

Source	Primary Task	Reported Effects	N
Michon, 1966	Experiment 1 Choice--reaction time	Yes	6
	Experiment 2 Maze--screw sorting, multiply, letter detection, Bourdon test	Yes	5
Brown et al., 1967	Car driving	No	8
Atkinson and Whitfield, 1972	Hovercraft maneuvering: drive the craft on a course	Yes	14
Vroon, 1973	Choice RT:		40
	- respond with same hand as with tapping	Yes	
	- respond with different hand than with tapping	No	
Vroon and Vroon, 1973	Choice RT:		40
	- predictable signal	Yes	
	- random signal	No	
Johnson et al., 1974	Visual signal detection	Yes	6
Johansen et al., 1976	Flight Simulator: Manual responses to auto-pilot failure	Yes	5
Shingledecker, 1980	Tracking	Yes	6
Casali and Wierwille, 1983	Flight simulator: respond verbally to auditory commands	No	39
Shingledecker et al., 1983	Tracking	Yes	4
	Memory search	No	10
	Visual monitoring	No	4
Casali and Wierwille, 1984	Flight simulator: manual responses to "danger" conditions	Yes	48

The literature review indicates that tapping variability increases when the tapping task is performed with a primary task that places a heavy burden on motor response generation. For example, Michon (1966) reported greater increases in PML for maze and screw sorting tasks relative to multiplication, letter detection, and the Bourdon test. In addition, increases in tapping variability have been shown with flight simulator (Johansen et al., 1976) and hovercraft maneuvering (Atkinson and Whitfield, 1972) where the responses to the primary task were manual. On the other hand, research by Casali and Wierwille (1982) with a flight simulator did not show increases in tapping variability in the dual task condition; however, this study involved verbal responses to auditory commands in the primary task.

Additional dual task research by Shingledecker (1980) and Shingledecker, Acton, and Crabtree (1983) supports the above contention that the tapping task is principally sensitive to concurrent tasks which place a burden on motor response generation. For example, Shingledecker (1980) found that tapping variability increased as a function of tracking difficulty. Furthermore, Shingledecker (1983) combined the tapping task with three different primary tasks: unstable tracking task, memory search, and a visual monitoring task. Tapping variability was shown to vary as a function of tracking difficulty but did not significantly vary in the memory search and visual monitoring tasks. The memory search task appears to tap resources associated with working memory processing and the visual monitoring task is associated with resources devoted to perceptual processing (e.g., processing of visual signals). These results are consistent with a multiple resource model (e.g., Wickens, 1981) since changes in tapping variability were only observed when the tapping task was performed with the unstable tracking task--a task which appears to tap resources principally associated with motor response processing.

Finally, related research by Vroon (1973) and Vroon and Vroon (1973) showed that tapping variability increased when subjects performed the tapping task and a choice reaction time task with the same hand; however, tapping performance was relatively stable when the tasks were performed with different hands. In addition, tapping variability increased in a task where the primary choice reaction time task involved predictable signals but not when

234

the signals were random. Vroon interpreted these results in terms of motor response expectancy. For example, tapping rate decreased shortly before stimulus presentation in the predictable signal condition but remained relatively stable in the random signal condition.

The above experiments indicate that performance on the tapping task (e.g., PML or tapping variability) is diagnostic of motor output loading. This was essentially the interpretation provided by Shingledecker et al. (1983) where the tapping task was paired with three different primary tasks. The present review of the Michon tapping task provides support for a multiple resource theory of information processing (e.g., Wickens, 1981). That is, tapping variability was not affected by primary tasks that utilized verbal responses (e.g., Casali and Wierwille, 1983), or which did not impose much of a burden on manual responding (e.g., the memory search and visual monitoring tasks in Shingledecker et al., 1983). Dual task decrements (as indicated by increases in tapping variability) are only evident when tapping is performed with primary tasks that impose heavy demands on motor response generation (e.g., maze task, screw sorting, and tracking).

RELIABILITY

Measures of reliability such as test-retest have not been determined for this task. However, Shingledecker (1984) reports that subjects reach a stable level of tapping performance after 15 minutes of practice. This task should be evaluated for test-retest reliability and stability of performance if it is to be used in repeated measures designs.

VALIDITY

The literature indicates that performance on the tapping task in a dual task condition is diagnostic of the motor output load imposed by the primary task. That is, this task is related to a general construct of motor response timing.

The tapping task was designed to be used as a secondary task. Therefore, measures of predictive or concurrent validity for the tapping task as a

235

stand alone task may not be meaningful. That is, correlating BTL with a
host of other performance measures may not be very fruitful since subjects
appear to be able to tap at a predetermined rate (e.g., one per second)
with very little practice. However, the above statements are based on a
few studies that did not explicitly investigate the predictive or concur-
rent validity of the tapping task.

SENSITIVITY

This task has shown sensitivity in dual task experiments to primary tasks
that impose demands on motor output performance. In addition, Johnson
et al. (1974) employed the tapping task (foot tapping) and visual signal
detection in a dual task combination to study the effects of carbon monox-
ide on performance. This study found an impairment in time sharing per-
formance as carboxyhemoglobin increased. This was especially true when the
signal detection task was demanding.

The above study represents the extent to which the Michon tapping task has
been utilized in behavioral toxicology research. In addition, the task has
not been employed in environmental stress or drug research.

TECHNICAL DESCRIPTION

A paddle shaped key (approximately one and one half inches by three inches)
which operates a microswitch is used to perform the tapping response. The
subject taps with the forefinger of the preferred hand. Intervals are
timed from the onset of one response to the onset of the next response.
Keybounce phenomena may be avoided in hardware or software design. In
addition, intervals of less than 10 msec should be rejected as spurious
input.

Trial Specifications

This test does not involve the presentation of a stimulus, rather the sub-
ject generates key taps based on a rhythm of 1- to 3-taps per second. Each
test period lasts 3 minutes and will consist of the following steps:

(a) a ready signal is presented on the CRT, (b) after the first tap, the screen clears and the message RESPONSES ARE BEING RECORDED is displayed, (c) the subject taps on the key at a steady rate for 3 minutes, (d) after the 3 minutes have elapsed the screen clears and the message TEST IS OVER is displayed. The above visual cueing signals can be replaced with auditory signals.

DATA SPECIFICATIONS

Unprocessed data for the task is a record of the duration in milliseconds of each successive tap. Summary statistics include two measures of tapping performance: the standard deviation of interval durations and the IPT variability score (see formula on page 230). Michon (1966) suggested the IPT variability score because it corrects for the partial dependence of error magnitude on interval duration. A lower IPT variability score indicates more temporally regular tapping and better performance. Typical IPT variability scores range from 10 to 40 (Shingledecker, 1984).

TRAINING REQUIREMENTS

Practice tapping for 15 minutes is adequate for training (Shingledecker, 1984). Subjects should be instructed to tap at a "personal rate" between one and three times per second, and to become as automatic as possible. Initially, six 30-second practice trials should be run to allow the subject to establish and maintain an acceptable tapping rate. The experimenter may need to coach the subject during these trials. It is best if a 2-taps per second rate is established early in training so that subsequent drift in tapping rate does not lead to unacceptable data. Four 3-minute trials should then be completed to provide sufficient practice, for a total of 15 minutes of training.

To summarize, the training phase for this test should consist of the following steps:

1.   Read instructions to the subjects.

237

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

The purpose of the Interval Production Task is to test your timing ability. To do this, we will have you tap a key at a constant rate. By repeatedly tapping the key you are producing time intervals between the taps. The more consistently you tap the key, the more equal will be the time intervals that you produce. Try to tap the key softly, but make sure that you press the key to the base on your taps. The best tapping rate is about 2-taps per second. We will do a few practice trials so that you can tell about how fast that is. The tapping task is run in 3-minute periods. You will be signalled at the beginning of the tapping period and again when the 3 minutes have past.

Section 21
STROOP TEST (UTC-PAB TEST NO. 20)
(INTERFERENCE SUSCEPTIBILITY TO RESPONSE COMPETITION)

PURPOSE

This test is a modified version of the classic color-word test developed by Stroop (1935). The purpose of this test is to measure a subject's susceptibility to response interference.

DESCRIPTION

During this test, both color and noncolor words are presented one at a time on a CRT screen. All words are displayed in the colors red, blue, or green and the subject is required to press one of three color coded keys that corresponds to the color in which the word is presented.

Three versions of this test are available for selection and are designed to produce different response time performance. The following represents a brief description of the three test versions: (a) the Control Version of this test contains three possible stimuli which are listed in Figure 18 under CWC (color-word congruent). This version of the test is intended to be used with the Interference Version; however, it may be used by itself as a choice reaction time task; (b) the Interference Version contains the six CWI (color-word incongruent) stimuli presented in Figure 18. This version represents the usual interference condition found in the Stroop color-word test; and (c) the Combined Version utilizes the six CWI and six NW (neutral words) stimuli presented in Figure 18. This version of the test represents the usual procedure that is employed in the examination of response interference. That is, stimuli that are relatively free of response interference (e.g., NW) are presented with those that produce maximum interference (e.g., CWI). The difference in reaction time between CWI and NW is indicative of response interference where such factors as stimulus encoding and response generation have been equated.

CWC STIMULI

BLUE$_b$ RED$_r$ GREEN$_g$

CWI STIMULI

BLUE$_r$ RED$_b$ GREEN$_r$

BLUE$_g$ RED$_g$ GREEN$_b$

NW STIMULI

DOOR$_r$ HOUSE$_r$

GUN$_b$ HOUSE$_b$

DOOR$_g$ GUN$_g$

Figure 18. Color-Word Stimulus Combinations for the Three Types of Stimuli [Note: The Lower Case Subscript Refers to the Ink Color (r = red, b = blue, and g = green)]

BACKGROUND

The original Stroop color-naming test (Stroop, 1935) required subjects to name a series of color patches that contained incongruent color words (e.g., the word "blue" in red ink). Relative to a control card (asterisks in color, color patches, or neutral words on color patches), the above card yielded much longer naming times. Jensen and Rohwer (1966) have provided an extensive review of the Stroop literature including methodology, research findings, and theoretical considerations. Much of their review deals with individual differences in performance as these relate to other performance and personality measures. On the other hand, the more current review by Dyer (1973) deals with experiments which were designed to extend knowledge of the Stroop phenomenon itself and experiments which utilize the Stroop phenomenon to study other problems such as word meaning, semantic satiation, and hemispheric differences.

The Stroop color-word test has been administered under two general paradigms: (a) a continuous procedure where subjects are presented cards with a series of color-words printed in incongruent ink colors (CWI), color-words printed in congruent ink colors (CWC), color blocks (CB), noncolor words printed in different colors (NW), or color-words printed in black ink (BW) and are required to read the words or name the colors as quickly and as accurately as they can; and (b) a discrete procedure where single stimuli (CWI, CWC, NW, CB, or BW) are presented for verbal or manual response. Procedure (b) has the advantage of providing discrete reaction times for each stimulus whereas procedure (a) results in a latency measure which is an aggregate over a series of responses. Furthermore, procedure (a) requires the careful construction of cards that control for such factors as the frequency of occurrence of each ink color and color-word per line, sequential repetitions of the same ink color or color-word, "suppress-say" (e.g., the word "blue" in red ink followed by the word "green" in blue ink) sequence, and "say-suppress" (e.g., the word "red" in blue ink followed by the word "blue" in green ink) sequences.

The Stroop test has yielded a variety of scoring procedures that fall into two general categories: (a) the basic time scores (e.g., CWI, CB, and BW),

and (b) derived scores based on the basic scores. The most frequently used derived scores are CWI-CB and CB-BW. According to Jensen (1965), the Stroop test contains three dimensions of variance. The three factors are referred to as Speed (SP), Color difficulty (Cd), and Interference (Int). Jensen (1965) argues that condition BW taps Sp; condition CB taps Sp + Cd; and condition CWI taps Sp + Cd + Int. Table 14 (adapted from Jensen and Rohwer, 1966) shows the intercorrelations between basic scores and two derived scores for 436 subjects. Note that the factors themselves (Sp, Cd, and Int) have very low intercorrelations and the large intercorrelations exist only between variables containing common factors. As can be seen, CB-BW is assumed to tap Cd and CWI-CB taps Int.

TABLE 14. INTERCORRELATIONS AMONG STROOP SCORES (N = 436)
ADAPTED FROM JENSEN AND ROHWER (1966)

Factors	Sp	Sp+Cd	Sp+Cd+Int	Cd	Int
Scores	BW	CB	CWI	CB-BW	CWI-CB
BW	--	.52	.43	-.07	.21
CB			.66	.82	.18
CWI				.48	.86
CB-BW					.06

There are two general hypotheses that have been proposed to account for color-word interference. The theories are the following: (a) response competition, response conflict, or output interference; and (b) perceptual encoding or input interference.

The most prominent explanation of color-word interference has been that of response competition or output interference (Dryer, 1973; Flowers, 1975; Keele, 1972; Posner and Boies, 1971). Briefly this theory states that when subjects are responding along a single dimension of a multidimensional stimulus (e.g., Stroop color-word test), both the relevant and irrelevant dimensions are automatically encoded. When the relevant attribute is ready for output, there are two or more (depending upon the number of dimensions) responses ready, and only one must be selected; responses to the relevant and irrelevant attributes compete for a single motor outlet (e.g., Klein, 1964; Morton, 1969). On the other hand, input interpretations of color-

word interference suggests that interference results from attempts to selectively attend to and process only relevant information (e.g., Treisman, 1969), or it results from a limited capacity for or the serial processing of information during input (Hock and Egeth, 1970).

Research employing physiological measures (e.g., average evoked responses) has supported the output interference hypothesis. Duncan-Johnson and Kopell (1981), using a discrete trials procedure of the Stroop task, found that response time varied with the congruence between the stimulus word and the color in which it was printed; however, the duration of stimulus processing, as indexed by P300 latency, remained constant. On the other hand, P300 latency was affected by the discriminability of the ink colors; that is, P300 latency increased as the ink colors were made less discriminable. There is convincing evidence that the latency of the P300 component of the human event-related brain potential reflects stimulus evaluation process that is independent of the time involved in response production (Pritchard, 1981). Therefore, the above results support the hypothesis that the Stroop effect (color-word interference) is primarily an output, rather than an input phenomenon.

RELIABILITY

Measures of reliability are not available for the present version of the test. However, Harbeson et al. (1982) report reliability data for conditions CWI, CB, BW, BW-CB, and CWI-CB. Their study involved a group testing procedure where subjects responded manually (i.e., pressed keys labeled with the first letter of the color names) to the meaning of the words or the color of the ink. The dependent measure was the number of words or colors correctly identified in a 30-second period (there were 100 color blocks or color-words per card arranged in a 10 by 10 matrix). The average performance (mean) for BW, CB, CWI, BW-CB, and CWI-CB were stable after six days of practice, while the variances were stable from the first day. The reliability coefficients for conditions BW, CB, and CWI were .81, for the derived scores BW-CB and CWI-CB were .22 and .23, respectively. Also Jensen (1965) reported reliability coefficients of .88 for BW, .79 for CB,

243

and .71 for CWI. Jensen's study involved verbal responses to the stimuli in the usual continuous paradigm (N = 436).

The proposed version of the Stroop test for the UTC-PAB menu differs procedurally from the above studies. For example, the UTC-PAB version of the test will employ discrete trials whereas the above versions used continuous paradigms. Therefore, the above reliability information may not apply directly to the UTC-PAB version of the test.

VALIDITY

Apart from its considerable face "validity," the assumption that this is a test measuring response competition (or conflict) is supported by behavioral research (e.g., Dyer, 1973; Flowers, 1975; Keele, 1972; Posner and Boies, 1971). In addition, research employing physiological measures has also supported a response interference interpretation of the Stroop effect (e.g., Duncan-Johnson and Kopell, 1981; Warren and Marsh, 1979).

The Stroop interference effect (CWI-CB) has been correlated with a wide variety of perceptual, memory, and intelligence tests. Jensen and Rohwer (1966) report that the Stroop interference factor has not been shown to significantly correlate with measures of intelligence; however, the interference factor has been shown to be significantly correlated with digit span (r = -.28) and serial learning of trigrams (r = .43). In addition, the interference factor has been shown to correlate significantly with performance on size estimation, rod and frame, embedded figures, and a field-dependence index (Gardner et al., 1959, cited in Jensen and Rohwer, 1966). However, the correlations were only statistically significant for the female subjects (r ranged from .37 to .67).

The above data indicate that the Stroop interference effect is related to a diverse set of other psychological variables, although nearly always quite low. This suggests that whatever processes are tapped by the Stroop test, they are of a very basic and broad significance.

SENSITIVITY

The Stroop test has been used extensively in the area of drug research.
Jensen and Rohwer (1966) report the results of a variety of studies which
indicate that stimulant drugs (e.g., methamphetamine, imipramine hydrochlo-
ride) improve performance (i.e., decrease the magnitude of the interference
effect), while depressants (e.g., amobarbital) and psychotomimetics (LSD)
have the opposite effect. Furthermore, nicotine has been shown to decrease
the interference effect (e.g., Wesnes and Warburton, 1978), while scopol-
amine and atropine increase it (e.g., Calloway and Band, 1958; Ostfeld and
Aruquete, 1962). Finally, the Stroop test has been shown to be sensitive
to age and psychiatric disturbance (Jensen and Rohwer, 1966).

TECHNICAL DESCRIPTION

The test will contain color words or noncolor words displayed in one of
three different colors: red, blue, and green. The stimuli will be pre-
sented one at a time on a CRT screen, subjects will classify the stimuli on
the basis of color by pressing one of three colored keys. The test will
contain three types of stimuli: (a) color words--red, blue, and green
printed in the color they name (CWC); (b) color words--red, blue, and green
printed in a color which does not match the meaning (CWI); and (c) neutral
words--gun, door, and house printed in red, blue, or green (NW). There
will be three stimuli for CWC, six for CWI, and six for NW. The stimuli for
these conditions were presented in Figure 15. The following is a descrip-
tion of the three versions of this test which will be available.

Control Condition (Version 1)

This condition will contain three possible stimuli (the three CWC stim-
uli). Each stimulus will be presented 12 times, yielding a total of
36 trials. The 36 stimuli will be presented in a random order.

Interference Condition (Version 2)

This condition will contain six possible stimuli (the six CWI stimuli).
Each stimulus will be presented six times, yielding a total of 36 trials.
The 36 stimuli will be presented in a random order.

Combined Condition (Version 3)

This condition will contain 12 possible stimuli (six CWI and six NW stim-
uli). Each CWI and NW stimulus will be presented six times. The 72 stim-
uli will be presented in random order.

Trial Specifications

For all conditions, the stimulus will remain on the screen until the sub-
ject makes a response. Immediately following the subject's response, the
screen will blank until the next trial. There will be a brief inter-
stimulus interval (ISI) following the conclusion of one trial and the
beginning of another trial. The length of this ISI will be randomly deter-
mined; however, it will fall within the limits of 1 to 3 seconds. If the
subject presses a response button during the ISI, the message "DO NOT PRESS
THE RESPONSE BUTTON UNTIL THE WORDS APPEAR" will be displayed for 5 sec-
onds. The stimulus will be presented on the screen such that it will be
centered both horizontally and vertically. The letters in the stimulus
word will all be in upper case and will be 1 inch tall. The response
manipulandum will be a box, separate from the keyboard, that has three but-
tons arranged in a horizontal row. One button will be colored red, one
button will be colored blue, and the remaining button will be colored
green. The buttons will be approximately 1 inch in diameter and will
require 3 to 7 ounces of pressure to depress. Response latency (the period
of time immediately following stimulus presentation up to the subject's
response) will be measured with less than 1 msec error.

DATA SPECIFICATIONS

For each trial the response latency will be recorded. The button pressed by the subject, the actual display color, and whether the trial was a CWC, CWI, or NW stimulus will be recorded for each trial. The following summary statistics will be provided for the response latencies; mean, median, range, and variance. In addition, the total number of correct responses will be determined. For the Control Condition the above statistics will be based on 36 trials employing CWC stimuli. For the Interference Condition the above statistics will be computed for the 36 CWI stimuli. Finally, in the Combined Condition the above statistics will be computed separately for the CWI and NW stimuli. Provisions will be made for the user to easily examine the individual trial data when desired. Provision will also be made for obtaining hardcopy printout of both the individual trial data and the summary data.

TRAINING REQUIREMENTS

The first phase of the test will consist of presenting the instructions to the subjects. The instructions are written so that they can apply to any of the three test versions. These instructions should be read to the subject before the start of the training trials.

Following the instructions, subjects should be presented with a minimum of 10 training trials (per test version). The nature of the training trials will depend upon the condition that is being run: (a) the Control Condition will involve 10 randomly selected CWC stimuli; (b) the Interference Condition will involve 10 randomly selected CWI stimuli; and (c) the Combined Condition will involve five NW and five CWI stimuli that are randomly chosen. If, on the training trials the subject presses the wrong response button, the message "PRESS THE BUTTON CORRESPONDING TO THE DISPLAY COLOR" will appear for 5 seconds. Following this message the same trial will be presented again.

The experimenter should carefully evaluate the subject's performance during the training trial to insure that the instructions are being followed. For

example, subjects should be reminded that they are to respond as quickly and as accurately as possible.

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

This is a test on the speed and accuracy of decision making. (Note: instructions in parenthesis apply to the combined version.) In this test you will be shown words printed in different ink colors. The words will be BLUE, RED, GREEN, (BLUE, RED, GREEN, DOOR, GUN, HOUSE) printed in one of the following colors; blue, red, or green. Your task will be to respond to the ink colors while ignoring the meaning of the words.

In this test, the words will be shown one at a time in the center of the CRT screen. Each trial will have the following steps: (a) a blank white field will be shown for about 1 to 3 seconds, and (b) a word printed in one of three colors will be presented. You are to respond to the stimulus by pressing the key with the color patch which matches the ink color of the stimulus. For example, if you were to see the word BLUE printed in red, you should quickly press the button with the red color patch. After you respond, the word "CORRECT" or "INCORRECT" will be displayed on the CRT for a brief moment. Following the feedback, the screen will clear and the

above sequence will be repeated (i.e., blank field, stimulus word, feedback).

For this test it is very important that you respond as quickly and as accurately as you can. The number of errors that you make and the speed with which you make your decisions will be recorded.

DICHOTIC LISTENING TASK (UTC-PAB TEST NO. 21)
(AUDITORY SELECTIVE ATTENTION)

PURPOSE

This test evaluates information processing resources dedicated to auditory
selective attention.

DESCRIPTION

Subjects are required to attend to a list of letters and digits that is
being presented to one ear while ignoring similar information being pre-
sented to the other ear. Subjects are to respond to the numbers presented
on the command ear channel by pressing corresponding number keys on a key-
pad in the order of their occurrence in the auditory message. Upon the
presentation of a specified auditory cue in the attended ear, the subject
either rapidly switches attention to the previously unattended ear or main-
tains attention to the previously attended ear, depending upon previous
instructions. Responding as per the current command ear is continued
throughout. The ear which is to be the command ear at the start of the
task is determined by the experimenter. The stimuli are produced by a
computer controlled speech synthesizer and are presented over dual channel
headphones.

BACKGROUND

Development of UTC-PAB Version of the Dichotic Listening Task

This task has been developed as a result of the importance of selective
attention resources in applied situations. For example, Gopher and
Kahneman (1971) point out that the failures of many flight cadets can be
traced to their inability to appropriately divide attention among concur-
rent signals. Gopher and Kahneman also assert that most studies dedicated
to the investigation of auditory selective attention utilize dichotic

250

listening (Broadbent, 1958: Moray, 1969; Neisser, 1967; Treisman, 1964, 1969).

However, the use of dichotic listening tasks in these investigations has not led to the standardization of tests using this method. In other words, according to Gopher and Kahneman (1971), the inconsistency among dichotic listening tasks has made it difficult for these studies to have significant impact on the problems concerning selective attention in applied settings. Thus, Gopher and Kahneman developed a dichotic listening procedure that attempts to provide information which can shed some light on the selective attention process as utilized in applied settings (e.g., flying of high performance aircraft). It is this procedure from which the UTC-PAB version of the dichotic listening paradigm was developed.

To summarize the specific paradigm of Gopher and Kahneman (1971): a series of 48 pairs of different messages is presented simultaneously to the two ears. The items presented to each ear are digits and unconnected words, and the rate of presentation is two items per second to each ear. One of the two messages is designated by a tone as relevant; the subject's task is to repeat immediately all digits in the relevant message. Part 1 of the message lasts 8 seconds, during which either two or four target digits are presented to the relevant ear. A second tone is then presented to indicate which ear is relevant in Part 2 of the message. On half of the occasions, the same ear is relevant in both Part 1 and Part 2. Either immediately after the reorientation tone or after the interpolation of one of two irrelevant items, three pairs of simultaneous digits are successfully presented to the two ears, and, as in Part 1, the subject's task is to report the three digits which have been presented to the relevant ear. Gopher and Kahneman utilized this procedure to obtain experimental results from two groups of subjects. The first group consisted of 100 cadets in flight school, early in their training while the second group consisted of 95 pilots on regular duty. These results provided considerable validation of the original expectations of Gopher and Kahneman; that is, performance on this task was found to be very predictive of the level of flight success achieved by each of these subjects (see section on Validity). Errors associated with this task can be classified as errors of omission (the lack of

a response when one is required) or errors of intrusion (the commission of
an inappropriate response). The subjects of Gopher and Kahneman were found
to commit many more errors of omission than errors of intrusion. Thus, it
was the omissions data which were incorporated into the analyses which
indicated a relationship between task performance and flight success, val-
idating the original experimental rationale of Gopher and Kahneman (see
section on Validity).

The Dichotic Listening Paradigm: An Overview

The dichotic listening paradigm (i.e., subjects are presented with a dif-
ferent stream of verbal information in each ear) was originally developed
by Cherry (1953) in an attempt to provide a degree of resolution to the
"serial versus parallel processing" issue. The results of the Cherry
experiment implied that the processing of information is predominantly
serial; in fact, Broadbent's (1958) well known single-channel "Bottleneck"
model of attention is largely based on the results of dichotic listening
studies such as those of Cherry from the 1950s. The paradigm utilized by
Cherry was as follows: subjects were fitted with headphones through which
two different streams, one to each ear, of verbal information were deliv-
ered. Subjects were asked to "shadow" (repeat the message aloud as it is
delivered) only one of the streams. Thus, attention is directed at one of
the messages and not at the other. The hypothesis is that evidence against
a serial processing model and for a parallel processing model would be pro-
vided if it is shown that semantic aspects of the nonattended channel were
processed.

The results obtained by Cherry (1953) supported the formulation of a serial
processing model. Subjects were unable to recall any aspects of the
meaning of the nonattended message. Cherry concluded that nonattended
material is not processed at a semantic level. This interpretation was
shared by Broadbent in the formulation of his model which proposed that the
devotion of attention to one specific source of information eliminates the
potential for the processing of other information. It was not long before
contradictory evidence began to appear, however.

Evidence for the existence of parallel processing was documented as early as 1959 when Moray found that subjects were aware of the presentation of their own name in the nonattended message. Apparently, then, information that receives little or no attention is, nevertheless, monitored for specifically targeted information (e.g., one's own name, a familiar name, a topic of interest). There must be at least a very temporary awareness of the semantic nature of nonattended material.

The work of Treisman (1960, 1964) provides further evidence of this premise. Treisman (1960) employed a dichotic listening paradigm in which the two messages were semantically similar. Again, subjects were instructed to attend to only one of the messages (ears). In this situation, subjects were found to inadvertently switch ears and shadow the nonattended message. Apparently, the brain monitors the meaning of nonattended material all along, and if this material is semantically well-fitted to the attended material, it is automatically introduced into awareness, disrupting a subject's ability to maintain performance as per his instructions.

Treisman (1964) provided further evidence for semantic processing of nonattended information. This study utilized a group of bilinguals as subjects. These two messages were, once again, semantically similar. However, they were in different languages. Subjects' performance was disrupted in a fashion similar to Treisman (1960). This demonstrates the salience of the semantic monitoring of nonattended material. The semantic nature of information can "trigger" it into an individual's awareness, even if the information is in a different language than the material which is being attended.

These findings obviously called for the development of attention models that differ greatly from that of Broadbent (1958). Such models were established by Treisman (1964) and Neisser (1967). These models describe attention as a parallel process rather than a predominantly serial process as per Broadbent (1958). To summarize these models: all streams of incoming information are constantly monitored. The individual actively selects the material which will receive his/her attention. Once a given stream of information is being attended, an individual may be relatively unaware of

other material, but the brain is, nevertheless, actively monitoring this material for salient, targeted information. From these parallel models of attention, then, arose the concept of selective attention; that is, individuals actively accept some inputs and reject others.

Much research has been devoted to the investigation of this selective attention process. Most findings conform to Treisman's original model which asserts that selection can operate on two general levels: (1) in terms of physical characteristics of the stimuli, and (2) in terms of the semantic nature of stimuli. The selection process required in the UTC-PAB dichotic listening task falls into the first category, as this type of selective attention activity resembles that which is required in many applied settings; that is, a specific discriminable (based on its physical characteristics) signal calls for a change of attentional and behavioral focus. Much of the research related to this process is irrelevant with reference to the development of the UTC-PAB version of the task. The characteristic of this process which seemed salient to Gopher and Kahneman (1971) in their development of the task is that performance is often characterized by substantial individual differences. It seems logical that the ability to quickly and accurately switch one's focus of attention would be a valuable skill involved in the flight of aircraft. This has been shown to be the case (Gopher and Kahneman, 1971; Gopher, 1982), and therein lies the practical value of the UTC-PAB version of the dichotic listening paradigm.

RELIABILITY

Reliability data on this task are not abundant. However, in their investigation of potential components of high level skill, Keele and Hawkins (1982) provide information which implies that the UTC-PAB version of dichotic listening is characterized by sufficient reliability. This piece of research was dedicated to the investigation of the performance of high level skills. Efficient utilization of selective attention is considered to be such a skill and, thus, performance measures on the Gopher and Kahneman (1971) task were obtained by Keele and Hawkins. Scores were also obtained for six other procedures that are also representative of "high

254

level skill." Intercorrelations were performed among these scores associated with the various tasks. Also included in this set of data were correlations between sessions of the same task; that is, reliability values were obtained for each of these performance scores. The reliability value associated with error scores on the dichotic listening task is very high (r = .92). However, because the assessment of task reliability was not the impetus of this study, this value must be viewed with caution.

In summary, though such data are scarce, there are indications that this task may be characterized by a sufficient, and possibly a very great degree of reliability. That is, only one study was found that addressed the issue of test-retest reliability (Keele and Hawkins, 1982), and this study was not specifically designed to investigate the reliability of the dichotic listening task. Additional studies that focus on the reliability of this test need to be conducted in order to provide conclusive evidence regarding test-retest reliability.

VALIDITY

As has been mentioned, the specific parameters of this task were developed by Gopher and Kahneman (1971) in response to their perception of selective attention as a vital component of flight success. Gopher and Kahneman (1971) have conducted an analysis to test the validity of this assertion. In other words, is performance on this task truly related to the subsequent success of a flight cadet?

To answer this question, Gopher and Kahneman conducted a follow-up study on the careers of the 100 cadets who had participated in the development of the task. The career progress of these cadets was divided into three categories: (1) 17 cadets were rejected during initial training on light aircraft, (2) 41 cadets were rejected early in training on jet aircraft, and (3) 42 cadets reached advanced training on jet aircraft. This three-point criterion was correlated with previously obtained performance measures on the dichotic listening task. Several significant correlations were found. Most notable was the correlation between this three-point flight criterion and number of omissions, which seems to indicate a high degree of

predictive validity associated with this task in terms of subsequent flight performance (r = .26, p < .01). This is especially true on Part 2 of the task (i.e., following the tone) where the occurrence of three omissions appears to be a good cut-off point with respect to the three-point of flight criterion. In fact, 76 percent of the candidates rejected during training on light aircraft, 56 percent of the candidates rejected early in jet training, and 24 percent of the candidates in the highest criterion category committed three or more omission errors in Part 2 of the task. It is apparent that this task represents an independent contribution to the prediction of success in flight training. This was reinforced more recently by Gopher (1982) in his investigation of several potential predictors of flight training success. This dichotic listening task proved to be the strongest predictive factor included in the investigation.

SENSITIVITY

Investigations of the sensitivity of dichotic listening performance have traditionally involved two general categories of variables: subject variables and stimulus variables. Dichotic listening tasks are not typically included in the study of environmental stressors, nor are they used often in dual and secondary task paradigms. This is due to the theoretical background from which this task was developed. As has been mentioned, selective attention is the underlying construct associated with this task. Two salient features of selective attention (as determined via the utilization of dichotic listening and various other paradigms) are a relatively high degree of variability among individual subjects, and a substantial degree of importance in terms of performance in many applied settings (e.g., flying of aircraft, driving a car). Thus, most studies involving dichotic listening tasks have focused on the following areas: (1) determining characteristics of individual subjects that help predict the efficiency of a given subject's utilization of selective attention, and (2) determining characteristics of stimulus presentation that enhance the effectiveness of selective attention resources.

Subject characteristics which are related to dichotic listening performance include psychopathological status (Bush, 1977; Hemsley and Richardson,

1980), auditory evoked potentials (Schwent, Snyder, and Hillyard, 1976), and performance on various other perceptual information processing tasks (Mihal and Barrett, 1976). Bush (1977) and Hemsley and Richardson (1980) have found strong relationships between dichotic listening performance and schizophrenia; that is, schizophrenic subjects perform significantly worse than normal subjects on dichotic listening tasks. In fact, Hemsley and Richardson report that the relationship between schizophrenia and performance on such tasks can be described as a continuum, as performance becomes progressively worse with the severity of the schizophrenic disorder. This finding is in accordance with the widely accepted notion that schizophrenia is characterized by the inability to distinguish relevant information from irrelevant information. Thus, dichotic listening paradigms are useful in the arena of psychopathology.

The cognitive capabilities of subjects are also related to dichotic listening performance. Research by Mihal and Barrett (1976) represents an attempt to formulate an information processing model of driver decision making. The validity of this model is not the central issue in this discussion, however. The salient feature of this research from the frame of reference adopted here is the set of intercorrelations among the cognitive tests employed in this study. Correlations between dichotic listening performance and performance associated with four other perceptual information processing tasks are highly significant in the positive direction. These four tasks are as follows: (1) a rod and frame task, (2) an embedded figures task, (3) a choice reaction time task, and (4) a complex reaction time task. Interestingly, all of these tasks are similar to dichotic listening in at least one respect: they all require some degree of efficiency with respect to selective attention resources. In all cases, subjects must at some point focus attention only on the relevant aspects of the stimuli if they are to perform well. This study showed that there were significant individual differences in performance of the dichotic listening and other tasks and, in addition, it also implied that such differences associated with many tasks probably share a common source; effective continuous attention allocation. Performance on any of these tasks is probably predictive of performance on any of the others. This knowledge could be

valuable in terms of selection of personnel for various tasks in applied settings.

Physiological characteristics are also related to dichotic listening performance. Schwent, Snyder, and Hillyard (1976) investigated the relationship between averaged auditory evoked potentials measured from the scalp and dichotic listening performance and found the amplitude of the N_1 component of the auditory evoked potential to be a reliable index of the distribution of selective attention between auditory channels (ears). The latency (following the stimulus) associated with the initial appearance of this component is noticeably variable across individuals. Perhaps this latency has some bearing on the eventual effectiveness of an individual's utilization of selective attention (Schwent et al., 1976).

Among the stimulus variables which have been found to affect dichotic listening performance are pitch (Schwent et al., 1976), localization (i.e., spatial separation; Schwent et al., 1976), semantic characteristics (Moray, 1959; Treisman, 1960, 1964), and linguistic characteristics (i.e., the language of a given message; Magiste, 1984; Treisman, 1964). There is a central point of commonality among all of these studies; that is, respective enhancements of performance based on the manipulation of each of these variables can be traced to one general principle. This principle is one of contrast. When a subject is presented with more than one auditory message, he/she will be able to more efficiently focus on the attended message if the attended message and/or the command cues are readily discriminable from the nonattended material either in terms of pitch, localization, semantic nature, and/or linguistic nature.

TECHNICAL DESCRIPTION

Parameters for a 36 trial UTC-PAB version of dichotic listening are as follows: two computer controlled speech synthesis devices are used, one for each auditory channel. Auditory stimuli are presented via dual channel headphones at 75 db/L Q (RE: 20 P). The duration of each individual stimulus (letter or digit) is 0.7 seconds; an entire trial required 26.8 seconds; and a block of 36 trials (preceded by six practice trials) takes

approximately 20 minutes. Each trial is divided into two parts. Part 1
consists of the presentation of letter and digit sequences to each ear.
Digits are never presented simultaneously to the two ears, and no digit is
repeated in either sequence. Any simultaneous presentations of stimuli to
the two ears consist of identical or dissimilar letters, or a letter to one
and a digit to the opposite ear. Part 2 of each trial is initiated by a
command indicating which message (right or left) is to be attended by the
subject. The rate of stimulus presentation is one letter or digit per
0.9 seconds. Three examples of a UTC-PAB dichotic listening trial are
depicted below:

(1)
_____

Part 1

           Left ear: R 8 N S M Y 2 G B 7 F L 6 R L 5

   "Right" (Channel to be attended command)

           Right ear: Y L 3 S R 4 F Z 9 X F 0 F N 1 L

_____

Part 2

           Left ear: B F 4 3 7 9

   "Left" (Channel to be attended command)

           Right ear: G L 1 5 6 2

_____

(2)
_____

Part 1

           Left ear: R 8 P N 2 0 R N Y 5 N 9 6 L 1 F

   "Right" (Channel to be attended command)

           Right ear: F G P 3 F 1 M 6 G L 5 8 X B M 4

_____

Part 2

           Left ear: B 6 6 N 1

   "Right" (Channel to be attended command)

           Right ear: F P 2 3 Y

_____

(3)

Part 1

Left ear: B 1 M N B F 5 S R 3 R 6 B 9 2 0

"Left" (Channel to be attended command)

Right ear: F X F 2 9 P 4 S N P R X B 6 G 7

Part 2

Left ear: 8 G X 4 F 1

"Right" (Channel to be attended command)

Right ear: 2 0 5 3 B S

Subjects are required to respond only to the numbers from the attended
channel by pressing corresponding numbered keys on a keypad.

DATA SPECIFICATIONS

Gopher and Kahneman (1971) utilized two measures of raw data: (1) number
of intrusion errors (reporting of inappropriate digits), and (2) number of
omission errors (failure to report the appropriate digit). The continued
utilization of these measures in future analyses would seem to be advanta-
geous due to their observed positive relationships with task reliability,
validity, and sensitivity. Because the construct under investigation is
selective attention, these measures are examined as follows: the effi-
ciency of resources devoted to selective attention can be evaluated by
comparing performance measures obtained during Part 1 with those obtained
during Part 2. The nature of any errors in Part 2 can also be of interest
with reference to the efficiency or lack of efficiency of attentional
resources. In fact, Gopher and Kahneman (1971) have found that errors in
Part 2 can often be attributed to one of three sources (Gopher and
Kahneman, 1971): (1) incomplete correct series; all responses are taken
from the appropriate message, but some omissions are present, (2) series of
mixed origin; some responses are appropriate, but some intrusion errors
exist which can be traced to the "nonattended" message, and (3) series
taken from incorrect ear; nearly all responses are errors of intrusion
which can be traced to the "nonattended" message. The relative frequencies

of occurrence of these three error sources can be provocative with reference to the allocation of selective attention resources.

Summary statistics such as means, maxima, minima, and standard deviations can be computed from these raw data.

TRAINING REQUIREMENTS

Subjects are told that this is a test of their ability to attend to a single message when a potentially distracting second message is present. They are then given the instructions and are stepped through the procedures inherent to these instructions. Following the presentation of the first two paragraphs of the instructions, subjects should be fitted into the headphones with the red tag going on the right ear. Then two practice trials should be performed. At this point, the experimenter should carefully evaluate the performance associated with these two trials to insure that the subject understands the task and is following the instructions. If so, the final four practice trials and the 36 experimental trials can be completed. The most important aspect of the instructions to be emphasized is that the subjects are to attend to the digits embedded in the attended message, and that "O" is not a "zero."

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

These are minimal training requirements for this task. Performance has usually stabilized following the six practice trials.

INSTRUCTIONS TO SUBJECTS

This task involves the simultaneous presentation of two series of letters and digits; one series is presented in each ear. Your task is to concentrate your attention on the letters and digits you hear in a particular ear and to record only the digits heard in that series. The ear you must concentrate on is called the "target ear" and will be clearly defined as "right" or "left" before each series begins.

To better familiarize yourself with the task, put on your headphone and listen to a practice trial. Listen for the command "right" or "left." Then, listen for the digits interspersed among the letters coming through that particular ear. The tape will begin momentarily.

The "right" or "left" command that you heard at the beginning of each series designated the ear you would have concentrated on during an actual test trial. Did you hear the digits embedded in the string of letters?

You will now actually perform practice trials 1 and 2. Press the numbered key on the keypad that corresponds to the digits you hear through the target ear. Remember to record only the digits you hear in the target ear and that "0" is not a zero. Let me repeat that "0" is not a zero.

Okay, try the first two practice trials. Afterward, we will discuss any problems you may have had.

Now, you will complete four more practice trials. After these are completed, immediately prepare for a regular test series of 36 trials. The entire testing process will take approximately 20 minutes. If you have no further questions, we will start now. Stand by.

Section 23
UNSTABLE TRACKING TASK (UTC-PAB TEST NO. 22)
(MANUAL RESPONSE CONTROL)

PURPOSE

This task tests information processing resources dedicated to the execution
of rapid and accurate manual responses.

DESCRIPTION

Subjects are required to view a video screen which displays a fixed target
area at the center. A cursor moves vertically from this target while the
operator attempts to keep the cursor centered over the target via rotary
movement of a control knob. The system is inherently unstable; operator
input introduces error which the system magnifies so that it is increas-
ingly necessary to respond to the velocity of the cursor movement as well
as cursor position. Based on two tracking performance measures (average
absolute tracking error and number of control losses) and a subjective
measure (task difficulty ratings), three reliably different demand levels
have been established by Shingledecker (1984) via systematically varying
the degree of instability in the system; that is, the rate at which the
cursor moves away from the target in rad/seconds. This value is repre-
sented by λ (Lambda).

BACKGROUND

This task was originally developed by Jex, McDonnell, and Phatak (1966).
Jex et al. (1966) point out that the more basic origins of this task came
about as a result of work in the analytical treatment of aircraft handling
qualities. Cited is the work of Ashkenas and McRuer (1959) who computed
just-controllable aircraft short-period static instability, and established
its strong relationship with operator (pilot) effective time delay. That
is, increased rate of system error associated with control tasks produces
corresponding increases in the operator's internal delay in processing and
responding to the disturbance. Subsequently, it was reported that control

263

loss occurred at the same static instability level for three test pilots (Jex and Cromwell, 1961). These findings resulted in a more extensive investigation of the measurement and dynamics of manual control behavior. The impetus for the development of a reliable, internally valid control task to be used in applied research settings had been provided. The main objectives of Jex et al. (1966) were, thus, to develop such a task and to experimentally validate the assumptions underlying a model of human control behavior.

Because tracking behavior involves input, translation, and output mechanisms, approaches to modeling such behavior have borrowed techniques from Fourier analysis and linear feedback control theory. Tracking performance can be described reasonably well by linear differential equations. Such equations are aptly called "transfer functions" and have been incorporated into a class of models referred to as quasilinear models of the human operator due to the fact that these models contain a linear component and a nonlinear component. Man's response to tracking input signals is nonlinear but it can, nevertheless, be approximated by a transfer function called the "describing function," plus the separate nonlinear component called "remnant." The value of the quasilinear approach stems from the fact that these models contain parameters that seem to correspond to specific characteristics of human control behavior in man-machine systems (e.g., time delay which reflects operator information processing, and gain which seems to reflect some higher level cognitive activity. Both will be discussed in more detail.)

A relevant example of such a model is the "crossover model" (McRuer and Jex, 1967) which employs a two-parameter (effective time delay and gain) describing function to model the proportion of the subject's response that is linearly correlated with the input signal (Figure 19, as depicted by Wickens, 1976, p. 3). As implied in the figure, this describing function takes the form $0(t) = K_s e(t - \tau_e)$, where $o(t)$ represents a subject's output at time (t), K_s represents a subject's gain, and $e(t - \tau_e)$ represents the input to the subject, or system error, seconds before. Thus, τ_e represents the subject's effective time delay; that is, the subject's internal delay in processing the tracking signal.
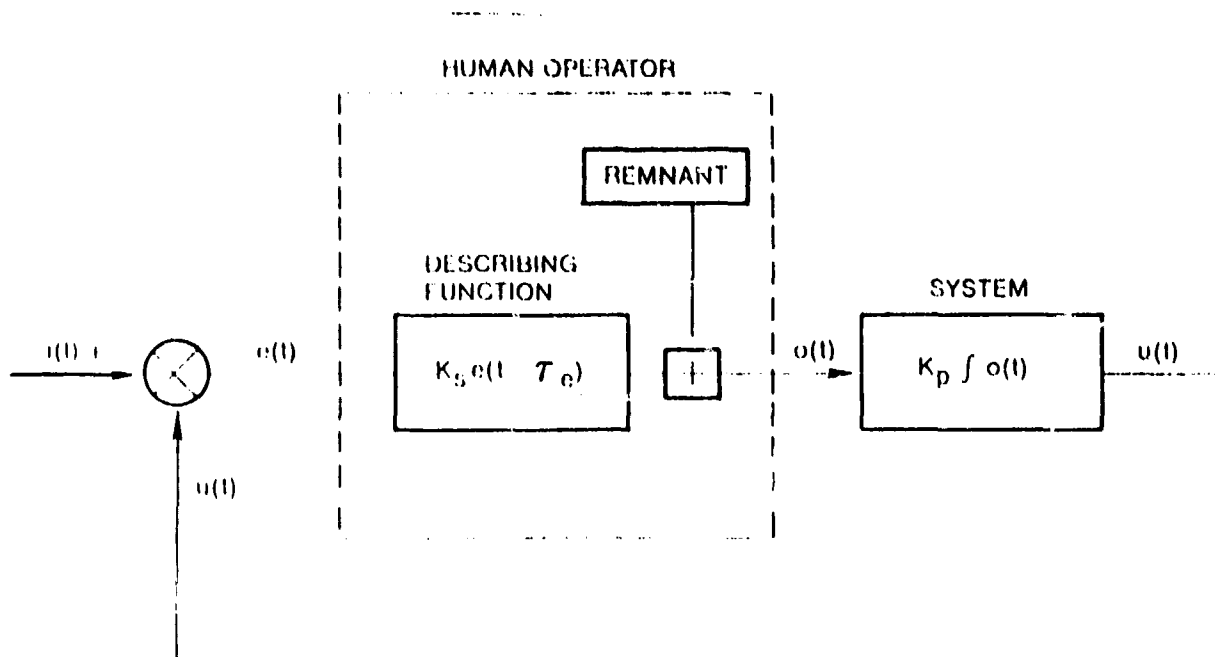
264

Figure 19. Block Diagram of Quasilinear Crossover Model

As has been mentioned, the effective time delay term measures the subject's internal delay in processing the tracking signal. This measure has been found to be somewhat analogous to discrete reaction time (Wickens, 1976); it is simply the time interval between the introduction of system error and the subject's emitting of an appropriate response to the error.

The gain parameter, K_s, is a measure of how large a corrective movement a subject will make in response to a given system error. Subjects who exhibit high K_s values tend to make relatively large amplitude control movements, leading to more oscillatory tracking behavior under some circumstances. Also, practiced subjects can adjust their gain to specified levels. In these respects, it can be said that perhaps gain represents something of a response bias, reflecting higher level cognitive processes (Wickens, 1976).

The key characteristic of the unstable tracking task is the positive feedback loop; that is, the inherent instability of the system. Once the system detects a control error, it will generate a proportional output error velocity whose value is determined by the gain. Unlike typical "purposeful" control in which this velocity is subtracted from the existing error by negative feedback, positive feedback adds the velocity to the error, increasing the rate of error movement away from the target. Wickens (1984) likens this to the dynamics of a balanced stick. If an error from the vertical is introduced, the stick will begin to fall, and the rate of falling (increase in error) will increase as it falls. In other words, within the positive feedback system, a subject's gain adds to the rate of system error. This is not true of negative feedback systems. It is especially integral to this task because it encourages subjects to make very precise, corrective movements.

While humans are better designed to deal with the properties of a negative feedback system, positive feedback loops are characteristic of many complex dynamic vehicles. These systems are potentially hazardous in that they necessarily require constant attention. For these reasons, it is important to understand the interrelationships of the elements of the describing functions associated with critical tracking behavior. And, the obvious

266

potential practical applications associated with this task render it a good candidate for utilization in dual task research and in the evaluation of environmental stressors and drugs on performance.

The UTC-PAB version of the unstable tracking task was designed with the following guidelines: (a) The unstable tracking task is based on a model of human information processing which posits three primary stages of processing and associated resources dedicated to perceptual input, central processing, and motor output or response activities (Shingledecker, 1984). The above model is based on multiple resource (Wickens, 1984) and processing stage (Sternberg, 1969) theories of human information processing. The unstable tracking task is assumed to largely tap motor output resources while minimally engaging perceptual input and central processing resources. An especially strong case can be made for this assumption since operator output directly influences the display. The operator is placing constant demands on motor output resources. (b) The actual nature of the present task was determined empirically in the test development phase by Shingledecker (1984). This research demonstrated that, based on two measures of tracking performance (average absolute tracking error and number of control losses) and subjective difficulty ratings, three reliably different demand levels are produced by lambda values of 1.0 (low demand), 3.0 (moderate demand), and 5.0 (high demand). Integrated tracking error scores and subjective ratings for these task conditions are presented graphically in Figure 20 (Shingledecker, 1984).

The fact that the task presents three increasingly difficult levels of task demand (associated with the three prescribed lambda values) has proved to make it especially amenable for dual task research. Shingledecker, Acton, and Crabtree (1983) evaluated the utility of performance on an interval production task (IPT) as a workload metric. Unstable tracking was one of the tasks employed in a dual task paradigm with the IPT. Three reliably different lambda values were employed to systematically manipulate task demand. The IPT did not interfere with tracking performance; that is, there were no significant differences from baseline tracking performance. However, there were systematic IPT variability increases associated with increases in tracking task demand. IPT scores were not affected by tasks
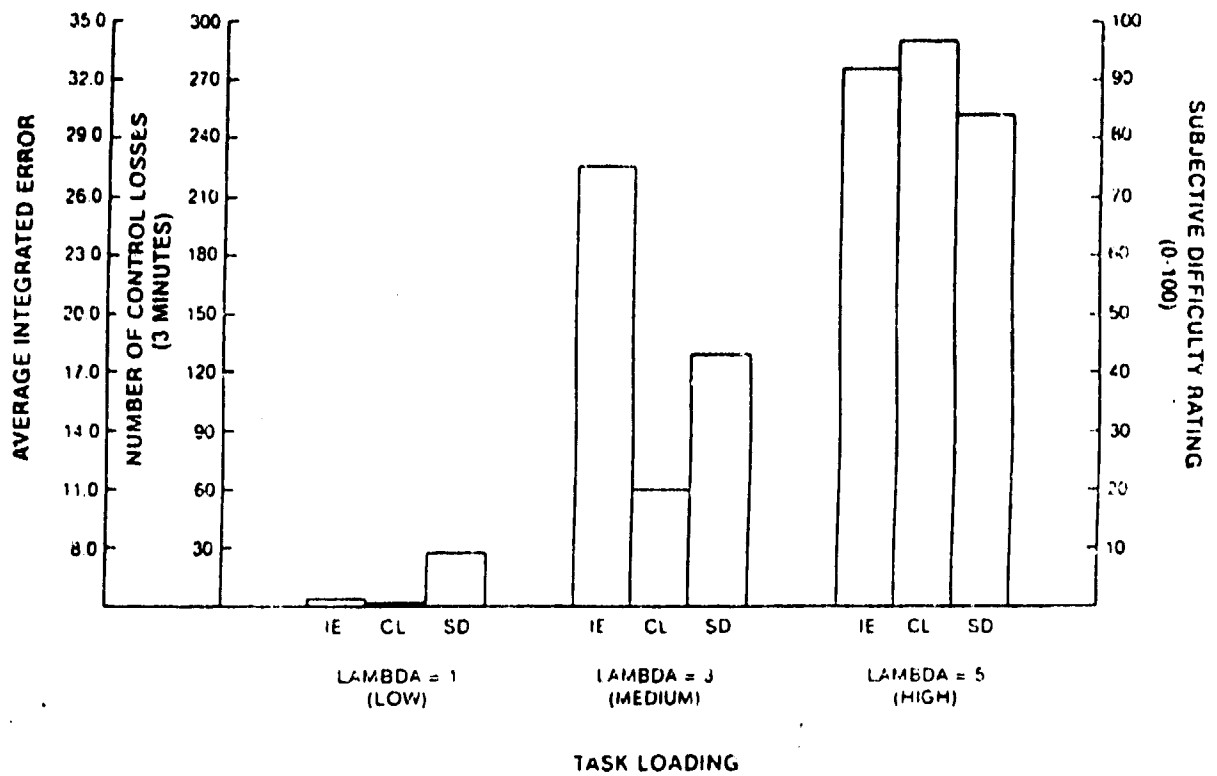
UNSTABLE TRACKING DATA

Figure 20. Unstable Tracking Data

which tap perceptual and central processing. Shingledecker et al. (1983) interpreted these findings as evidence that the unstable tracking task and the IPT place demands on resources devoted to motor responses and are not significantly related to perceptual or central processing. These findings are consistent with the multiple resource model of Wickens (1984).

RELIABILITY

The reliability and stability of critical tracking tasks are dependent upon the effects of practice (Damos et al., 1981; Damos et al., 1984). Damos et al. (1981) present test-retest reliabilities (intercorrelations) of mean critical tracking scores (the average degree of instability when control is lost). The correlations exhibit differential stability subsequent to session 10 (of 15). The mean r-value (n = 12) based on the final five sessions is .764, which is classified as moderate. Damos et al. (1984) also presented cross-session product-moment correlations of tracking performance based on critical λ scores. Again, performance stabilizes after 105 brief practice trials. The authors point out that although this is not considered to be an extensive or tedious practice period, it does represent more practice than is often utilized in studies that typically employ a tracking task (e.g., dual task, environmental stress evaluation). Performance from day 8 through 14 (the final day) shows slow linear improvement. Perhaps this would continue after day 14. The implications are that the task is sufficiently reliable for inclusion in dual task, environmental stress, or drug related research if proper attention is given to the importance of practice. That is, practiced subjects' performance is reliable, and any decrement could safely be attributed to the research setting. No reliability data based on average error or number of control lapses per trial have been located. (Note: differential stability is characterized by high, stable test-retest correlations.)

VALIDITY

In their development of the task Jex, McDonnell, and Phatak (1966) conclude that there is "good experimental validation of the theoretical assumptions and implications of the operator's behavior (with respect to the elements

269

of a describing function) in the first order critical task" (p. 142). The
experimenters arrive at this conclusion based upon their gathering of data
to establish an operator describing function. The three parameter
"Extended Crossover Model" of McRuer et al. (1965) was used to fit the
data. The form of this describing function is as follows:

$$Y_p(jw) = \frac{c}{e} = K_{pe}^{-j(\alpha/w + w \tau_e)}$$

where K_p = Gain

τ_e = Effective time delay

α Accounts for mid-frequency effects of the low frequency phase
droop (this parameter is not relevant to a discussion of the
human operator).

The data indicate that the τ_e level approaches an irreducible minimum and
flattens out as extreme instability (system error) is reached (see Jex
et al., 1966, Figure 4A). Also, experimental gain margins are found to
decrease as instability increases. Actual operator gain closely follows
the theoretical gain for maximum gain margin as delineated by the function;
gain limitations are constrained as critical limits are approached. All of
these findings are in very good accordance with the extended crossover
model. This experimentation represents good validation of the theoretical
implications of increased instability (λ) on the elements of the describ-
ing function (τ_e, K_p) which represent information processing resources
associated with the subject's production of manual control responses.

SENSITIVITY

Studies by Klein and Jex (1975) and Dott and McKelvy (1977) both show
tracking performance decrements associated with alcohol consumption. Klein
and Jex point out that traditional negative feedback tracking tasks have
shown little sensitivity to the effects of alcohol. However, the inherent
instability of the Critical Tracking Task (CTT) employed by Klein and Jex,
which is essentially the same as the UTC-PAB version of the unstable track-
ing task, is characterized by significant impairments with increases in a
subject's blood alcohol concentration. Dott and McKelvy also investigated
the sensitivity of an unstable tracking task to alcohol. Mean error, total
error, and the degree of instability when control is lost were measured.

All three performance measures showed significant decrements as a function of blood alcohol level (i.e., mean error and total error increased; degree of instability when control is lost decreased).

The sensitivity of unstable tracking to a secondary task(s) was examined by Wickens (1976) and Damos et al. (1981). Wickens employed two secondary tasks: (1) auditory signal detection, and (2) application of a constant force. The former represents an "input task" while the latter represents an "output task." The auditory detection task required subjects respond to 300 msec tones in a white noise background. These signal tones were pure sine waves at 1000 Hz. Tone intensity ranged from 59 db SPL to 63 db SPL. The subjects responded to the tones vocally, triggering a voice key. Response and signal occurrences were recorded for analysis. In the force application task, subjects grasped a vertically mounted isometric, force-sensitive control. Prior to trials which involved the force application task, subjects utilized visual feedback from a voltmeter to provide sufficient force to center the needle on the voltmeter. The visual feedback was terminated at the beginning of each trial and subjects then attempted to maintain this force for the duration of the trial. Wickens concluded that attentional limitations associated with the unstable tracking/secondary task paradigm are more severe for output than for input processing stages, as two of the three performance measures evaluated (Figure 21) were sensitive to time sharing conditions which involved the force application task. No such sensitivity was found with auditory signal detection. The fact that the tracking task interferes with the "output task" and not the "input task" can be interpreted as further support for the assumption that tracking essentially taps motor output resources. The dual task paradigm employed by Damos et al. (1981) required the simultaneous performance of two identical unstable tracking tasks. That is, two displays were shown side by side on a CRT. The right hand must respond to the right display, the left hand to the left display. The study evaluated the results in terms of implications concerning the concept of a "general time sharing ability." It was reported that dual task performance reached approximately the same level as single task performance after 15 sessions

271

(1) **Effective Time Delay** (Jex, McDonnell, and Phatak, 1966).

(2) **Operator Gain** (Jex et al., 1966).

(3) **Tracking Error** (Jex et al., 1966).

    (a) Mean Squared Error (Wickens, 1976).

    (b) Integrated Absolute Error (Adler, Strasser, and Muller-Limmroth, 1976).

(4) **Critical Track Score** (Damos, Bittner, Kennedy, and Harbeson, 1981; Damos, Bittner, Kennedy, Harbeson, and Krause, 1984).

(Note: A critical tracking score is the value of λ [the degree of instability of the controlled element] at which the operator can just control the system. This measure should reflect time delays associated with an operator's perceptual processing, neural transport, and neuromuscular systems as well as effective time delay of the display associated with a given value of λ.)

(5) **Dott and McKelvy (1977)** Table 1

    (a) t = total time (sec) from start of trial until control is lost.

    (b) tH = time (sec) while the rate of change of $F^* = 1.0$ rad/sec^2.

    (c) tl = time (sec) while the rate of change of $F = .25$ rad/sec^2.

    (d) T = total error score.

    (e) tH = error score during tH.

    (f) tL = error score during tL.

    (g) Fs = value F^*(rad/sec) when the rate of change of F transitions from 1.0 rad to .25 rad/sec^2.

    (h) Value of F (rad/sec) when control was lost

        $^*F$ = instability in the loop for which subject must compensate (in rad/sec); usually designated as λ.

Figure 21. Performance Measures--Unstable Tracking

of dual task practice. Perhaps dual task decrements in unstable tracking performance can be reduced or alleviated via extended practice.

Tracking tasks have frequently been employed in the study of the effects of acceleration (G-stress). Such research is of great practical significance as tracking behavior is involved in the control of an aircraft, and pilots frequently are exposed to G-forces. A great deal of this research has been done at the Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio. There is a considerable volume of such research, employing a wide range of tracking tasks, levels of G-stress, and other variables of particular interest to a given study. To briefly summarize the findings of G-stress/tracking research: tracking performance is generally impaired by exposure to G-forces; the magnitude of such effects can be influenced by the exact dynamics of the task and other variables often employed in such studies (i.e., direction of acceleration, subject position, G-force protective suits, etc., see reviews by Grether, 1971; Little, Hartman, and Leverett, 1968; Van Patten, 1984).

Jex, Peters, DiMarco, and Allen (1974) hypothesized that physiological deconditioning from orbital living (in the form of 10 days of enforced bedrest) could have potentially deleterious effects on a pilot's ability to control his aircraft manually in a shuttle reentry simulation. Subjects were provided with G-suits which protect them from the effects of G-stress. While this bedrest had no effect on mean critical scores (see Figure 20), a bedrest by centrifugation interaction was suggested. Before bedrest, subjects' (N = 42) critical scores were slightly better, though not significantly better (G-suits compensate for decrements, but do not enhance performance following a centrifuge run as compared to prerun). After bedrest, 62 percent of the postrun scores worsened relative to prerun scores. The enforced bedrest seems to interfere with G-protected subjects' ability to overcome the deleterious effects of G-stress.

Research by Adler, Strasser, and Muller-Limmroth (1976) showed that integrated absolute tracking error can be significantly lessened under conditions of distributed, as opposed to massed practice and monetary incentive. Also, a change in practice regime was found to produce deleterious

273

effects. These results imply that traditional models of control behavior should be modifiable with consideration to such "often ignored" variables as motivation, fatigue, learning, etc. (Note: The task utilized by Adler et al. (1976) is not the critical tracking task developed by Jex et al. (1966), but the two are comparable in many respects.)

In summary, positive feedback tracking is generally more sensitive to environmental stressors than negative feedback tracking. As noted by Klein and Jex (1976), alcohol had shown little effect on negative feedback tracking. As a result, these tracking tasks were not often employed in drug related research. The sensitivity of positive feedback tracking to alcohol effects has created an interest in the inclusion of this task in drug research. Secobarbitol and carbon monoxide are two substances whose effects on positive feedback tracking are very similar to those of alcohol (Putz, 1976). This can probably be attributed to the demands placed on motor control resources by the unstable tracking task, which are greater than the demands exerted on these resources by negative feedback tracking.

TECHNICAL DESCRIPTION

The unstable plant dynamics of the task are a first-order divergent element of the form:

$$P_{(s)} = \frac{\lambda\, e^{-ts}}{s - \lambda}$$

where λ (lambda) is selected by the experimenter to vary the task difficulty. The system display time delay term (t) in the above equation was not explicitly specified to be part of the desired dynamics, but is present in any digital implementation of a tracking loop. The magnitude of this delay was determined analytically to be no greater than 49 msec. It includes the 21-msec time frame (1000 msec/47 Hz), an 11-msec sample-and-hold (0.5 time frame) associated with display generation, and a 17-msec sample-and-hold associated with the television time frame (Shingledecker, 1984).

The real-time tracking loop software is free running (i.e., the iteration
rate is not directly controlled by clock interrupts). As a result, the
full 21-msec time frame is used for computation of the new cursor position
given the sampled stick value. Despite the fact that the tracking loop is
free running, the iteration rate (and accordingly, the time frame and trial
length) varies by less than 3 percent within or across trials. A trial is
flagged as invalid if the slight variations associated with these system
dynamics result in a trial length which varies by more than 5 percent from
the prescribed 3 minutes.

No external forcing function is applied to the tracking loop. The unstable
dynamics are simply excited by human tracking remnant and by noise in the
stick digitization process. If the subject loses control and the cursor
travel reaches the edge of the display, it is automatically reset to dis-
play center and the subject continues tracking. The active area of the
display is ±9.5 cm and the number of control losses is based on the sampled
value of each time frame. The software permits the user to break the trial
up into 1 second segments for detailed analysis of tracking performance.
Thus, at the finest level of resolution, the average absolute error scores
are based on 47 samples of instantaneous error (Shingledecker, 1984).

Calculation of the average absolute tracking error:

$$E = \frac{\sum_{1}^{n} e_i}{n}$$

where: e_i = absolute error in rad/second for a given
            time interval i.

         n = total number of time intervals
             utilized in analysis.

The cursor is intended to have the appearance of an aircraft viewed from
the rear and the target is a line segment drawn horizontal to the movement
line of the cursor.

DATA SPECIFICATIONS

Unprocessed data records will include average error scores for each con-
secutive 1 second interval of a 3-minute trial. Summary statistics will be
the average error score for the complete trial and a tabulation of the num-
ber of times the cursor leaves the extreme edges of the screen. (Note:
reliability data presented are based solely on critical scores. It is not
possible to obtain this measure with the UTC-PAB version of this task
because lambda is constant within a block of trials to exert a prescribed
demand level on manual output resources. See Figure 18 for a complete list
of potential performance measures.)

TRAINING REQUIREMENTS

All trials at any of the three loading levels are 3 minutes long. Instruc-
tions specify that the cursor should be kept centered over the target for
as much of the time as possible and that allowing the cursor to leave the
edge of the screen should be avoided. Subjects are given 10 seconds to
gain control of the cursor before the trial begins for data collection.
Major training (practice) effects are eliminated with six practice trials
at each loading level (Shingledecker, 1984). However, 10 to 12 practice
trials should be employed to enhance performance stability (Damos et al.,
1981; Shingledecker, 1984).

To summarize, the training phase for this test should consist of the
following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that
   the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require
   additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

The object of the Unstable Tracking Task is to keep a cursor centered over a target area in the middle of the screen of a CRT. You can control the movement of the cursor by turning the control knob. Rotating the knob to the right (clockwise) moves the cursor up, and rotating it to the left (counterclockwise) moves it down. The cursor appears at the center of the screen and naturally tends to move vertically away from the center. Try to keep the cursor centered over the target at all times. If the cursor reaches the edge of the screen, it will reappear at the target and begin moving away again. This is called a control loss and should be avoided if possible.

The task is run in 3-minute periods of data collection, called trials. The difficulty of the control task will vary from trial to trial. During some trials, the cursor will be fairly easily kept in the middle of the screen, but others will be more unstable. To start the task, rotate the control knob until the numerical display on the screen reaches zero. The task automatically stops after 3 minutes and the screen will go blank.

MEMORY SEAKCH-TRACKING COMBINATION (UTC-PAB TEST NO. 23)
(TIME SHAKING ABILITY)

PURPOSE

This dual task combination is intended to tap information processing
resources dedicated to time sharing ability; that is, the ability to per-
form two tasks concurrently.

DESCRIPTION

This is a dual task paradigm involving unstable tracking (UTC-PAB Test
No. 22) and the Sternberg Memory Search Task (UTC-PAB Test No. 9) as
employed by Wickens and Sandry (1982). Subjects are required to track with
their left hand and respond to the memory search stimuli with their right
hand. Stimulus and response parameters are as described for the single
task conditions in Sections 10 and 23.

To start a trial, the subject is shown the positive set for the Sternberg
task, as under single task conditions. This display is erased and the
trial begins 2 seconds later. Subjects are told to respond as quickly and
accurately as possible, and that both tasks are equally important.

BACKGROUND

Combinations of a memory search task and a tracking task have been employed
in research aimed at testing assumptions underlying multiple resource
models of attention. Also, this task combination has been employed to test
hypotheses regarding task-hemispheric integrity. The above areas of
research will be discussed in order to provide background information for
the UTC-PAB version of the task.

Research by Vidulich and Wickens (1981) employed a combination of a
tracking task with a memory search task. The memory search task was pre-
sented either visually or auditorially and responded to either verbally or

manually. Previous research has indicated that some mappings of input/output channels on tasks requiring a particular type of central processing are more efficient than others (Greenewald, 1979). Also, Wickens, Vidulich, Sandry, and Schiflett (1981) have argued that a unique compatibility relationship exists when verbal tasks are assigned to the auditory/speech modes, and spatial tasks to visual/manual modes.

The following results from Vidulich and Wickens (1981) are relevant to the discussion of the UTC-PAB version of the memory search-tracking combination. First, a verbal memory search task was performed best in the auditory input and speech response mode and most poorly in the visual input and manual output mode. This finding was consistent for both the single and dual task combinations. Second, tracking difficulty exerted a negligible effect on the memory search task when the input/output modalities of the two tasks were separate. This finding was expected since the central processing codes of the two tasks are also separate (e.g., verbal for the memory search task and spatial for the tracking task). Finally, the effect of visual input competition was borne mostly by the perceptual/cognitive memory search task, while the effect of manual output competition was observed in the response-loading tracking task.

Research by Schingledecker, Acton, and Crabtree (1983) also indicates that the memory search task is a perceptual/cognitive task, whereas, the tracking task places a heavy burden on response processing. In this study, the Michon tapping task (UTC-PAB Test No. 19) was paired with either a tracking task or a memory search task (a visual probability monitoring task was also used). The Michon tapping task was shown to interfere with the tracking task but not the memory search task. The Michon task is assumed to principally tap resources associated with response timing (see UTC-PAB Test No. 19 for a review of the tapping task) and, therefore, should not interfere with a task that does not place heavy demands on this resource. This differential result, in terms of dual task performance, supports the hypothesis that the UTC-PAB version of the unstable tracking task places a heavy burden on resources associated with response processing.

The UTC-PAB memory search-tracking task presents two different task con-
figurations that can be selected. The memory search task can be presented
either visually or auditorially. The above research indicates that the
auditory memory search task will be more efficiently time shared with the
tracking task than will the visually presented version. However, this ver-
sion of the task results in a combination where the two tasks share output
modalities (e.g., both tasks require manual responses) such that perform-
ance on the tracking task will be disrupted by the requirements to respond
to the memory search task. The tracking task is a continuous task with a
relatively heavy response component which can be disrupted by competition
for output resources. On the other hand, the memory search task is primar-
ily a perceptual/cognitive task which briefly demands output resources only
occasionally.

Research on task-hemispheric integrity in dual task performance (Wickens
and Sandry. 1982; Wickens, Sandry, and Hightower, 1982) is also relevant to
the discussion of the UTC-PAB dual task test. Task-hemispheric integrity
refers to a situation under dual task performance where the central pro-
cessing and response components of each task are associated exclusively
with a given cerebral hemisphere. For example, task-hemispheric integrity
should be achieved when a spatial task is performed with the left hand and
a verbal task with the right hand (Wickens, 1981). That is, the spatial
task is assumed to be processed in the right hemisphere and, therefore, if
responded to with the left hand, central processing and response processing
would be associated with the same hemisphere. A similar argument can be
presented for the verbal task which is presumed to be processed in the left
hemisphere.

Wickens and Sandry (1982) used two different versions of the memory search
task (e.g., a verbal and spatial variant of the task) in dual task combina-
tions with a tracking task. The results of the study indicated that
responding to the verbal memory search task with the right hand (integral
combination) resulted in greater time sharing efficiency relative to the
condition where the memory search task was performed with the left hand
(nonintegral combination). The results of the study also suggested that

280

the spatial memory search task and the tracking task competed for similar resources and, therefore, an "integrity" benefit could not be realized.

The initially proposed version of the memory search-tracking combination task in UTC-PAB presented the recommendation that the memory search task will be responded to with the right hand and the tracking task with the left hand. The reason for this response hand assignment is to obtain task-hemispheric integrity in this dual task combination. The proposed response hand assignment should be the one that results in the highest degree of time sharing efficiency based on the hemispheric integrity hypothesis.

To summarize, the UTC-PAB memory search-tracking combination task represents the combination of two tasks that compete for different pools of resources (e.g., perceptual/cognitive versus response--see UTC-PAB Sections 9 and 22 for reviews on the memory search and tracking tasks). In addition, the auditory version of the memory search task should be time shared more efficiently with the tracking task than the visually presented version (Vidulich and Wickens, 1981). The recommended response hand assignment should result in task-hemispheric integrity (Wickens and Sandry, 1982), thus, leading to relatively high time sharing efficiency.

The above research illustrates the uses of dual task methodology to test assumptions regarding human information processing (e.g., testing different theories). However, the UTC-PAB dual task combination will be used to test the effects of chemical defense treatment and pretreatment drugs. The reason for using a dual task combination in this context is to determine the effects of drugs on complex human performance. The memory search-tracking task combination has not been used in the above context. However, dual task methodology has been employed in the study of the effects of chemical and environmental stressors on human performance. For example, Putz and his associates (Putz-Anderson, Setzer, and Croxton, 1981; Putz, 1979; Putz, Johnson, and Setzer, 1974) have examined the effects of toxic substance on the performance of a tracking-tone detection task combination. This research has generally found a significant effect of stressor (e.g., carbon monoxide and alcohol) on tracking performance but not on the tone detection task.

Research by Houghton, McBride, and Hannah (1985) provides another example
of the uses of multiple tasks in the evaluation of environmental stressors
(e.g., G-stress induced loss of consciousness). Houghton et al. (1985)
used a multiple task arrangement consisting of: (a) two choice reation
time; (b) mental arithmetic; and (c) a two dimensional compensatory
tracking task. In this study, the above tasks were performed simulta-
neously where the tracking task served as the primary task and the others
were secondary tasks. The results, with respect to the effects of G-stress
induced loss of consciousness on complex performance indicated: (a)
significant impairment in the choice reaction time task and the mental math
task; and (b) there was no impairment in the primary tracking task.

The above studies show how dual task methodology can be used in the eval-
uation of complex performance under an environmental stressor. These
researchers employed dual task methodology as a means to create a complex
performance task with high processing load and some degree of relevance to
the operational environment. The UTC-PAB memory search-tracking combina-
tion appears to be a good candidate for the evaluation of stressor effects
on complex performance: (a) the combination of these two tasks result in a
test that taps a wide range of processing resources; (b) test difficulty
can be varied by increasing tracking and memory search difficulty, and (c)
it can examine, to a degree, the effect of drugs on a subject's ability to
efficiently time share.

RELIABILITY

The concept of task reliabilty is central to the evaluation of environ-
mental stressors since studies typically utilize repeated measures designs.
Research of this type usually involves the collection of data under base-
line and "treatment" (stressor) conditions for the purpose of comparison.
For this comparison to be meaningful, there is a requirement that the
repeated data collection under baseline conditions would yield very similar
(reliable) results. Unfortunately, there is no research that has assessed
the test-retest reliability of this dual task combination.

However, some evidence would lead one to believe that this combination is probably characterized by sufficient reliability. As has been mentioned in the discussion of the memory search and tracking tests, single task performance associated with each of these tests tends to be reliable. Tracking performance, in terms of critical instability scores, becomes stable after eight practice sessions, and there is a significant degree of reliability among scores from sessions 9 to 15 (Damos et al., 1981). In addition, Carter et al. (1980) found reaction times associated with the memory search task to be reliable after four practice sessions. Finally, the observed test-retest reliability of other dual task combinations involving tracking (Wickens, Mountford, and Schreiner, 1980) suggests that this combination may also be reliable.

However, simple test-retest reliability carries little weight when compared to a full investigation of task reliability carried out over 10 to 15 sessions as per Damos et al. (1981) and Carter et al. (1980). Such a study involving the tracking-Sternberg task combination would be required to draw any robust conclusions concerning task reliability.

VALIDITY

The findings of Wickens and Sandry (1982) can be interpreted to indicate that relative performance on this task combination is an index of one's ability to time share, since it was found that extensive practice can practially extinguish any single-dual task performance differences associated with this combination. The alternative interpretation, however, is that this sharing is made possible by the fact that these two tasks tap into two distinct pools of information processing resources. A subject can dedicate central processing resources (working memory) toward the memory search task and motor output resources to the tracking task. Whether or not there are resources specifically devoted to time sharing is not clear. Researchers have attempted to uncover a general time sharing factor, but the evidence is inconclusive (Wickens, Mountford, and Schreiner, 1980; Sverko, 1977). In summary, this task combination can be recommended for inclusion in studies attempting to assess time sharing abilty, with the provision that alternative interpretations of any results are borne in mind. Additional

283

research is required to help clarify this somewhat cloudy issue of construct validity associated with the tracking-memory search task combination.

SENSITIVITY

The relatively few investigations of the sensitivity of a tracking-memory search dual task combination have shown this combination to be sensitive to several variations of stimulus and response parameters (e.g., the order of the tracking task, the positive set size associated with the memory search task and/or which hand to use when responding to a given task). The respective rationales for such manipulations are rooted in the attempted assessment of multiple resource frameworks of information processing and/or hemispheric integrity (as mentioned earlier). As this task combination typically has been utilized only in studies such as these, little or no research has yet been performed which attempts to evaluate the potential effects of environmental stressors on tracking-memeory search dual task performance. However, this sensitivity to variations of task parameters serves as a preliminary indication that performance on the tracking-memory search combination could also be potentially sensitive to environmental stressors.

There is additional evidence which suggests that the tracking-memory search combination could be sensitive to environmental effects. Advantages (as compared to single task performance) in terms of task sensitivity have been attributed to other dual task combinations such as the tracking-choice reaction time combination employed by Putz (1979). Thus, perhaps performance associated with this tracking-memory search dual task combination could follow the same pattern and exhibit greater sensitivity to environmental stressors than single task, unstable tracking and/or single task, memory search, both of which have been found to exhibit an adequate degree of sensitivity to stressors (see the sections in this report for the unstable tracking and memory search tests).

TECHNICAL DESCRIPTION

Stimulus and response parameters are as delineated in the single task paradigms. In the memory search task, the numbers comprising the positive set are presented simultaneously for a duration of 1.5 seconds per item. Memory search stimuli are to the left of the tracking stimuli on the CRT. Response equipment is the same as under the single task conditions. The subject is shown the positive set of the Sternberg task to start the trial. The trial begins 2 seconds after the set is erased. Each trial lasts 90 seconds, and there is a 30-second break between trials.

DATA SPECIFICATIONS

Raw data collected are average root mean square (RMS) error (Unstable Tracking), percent error (Memory Search), average correct reaction time (Sternberg), and average incorrect reaction time (Sternberg). Standard summary statistics are the means and standard deviations (overall or per trial) associated with each dependent measure.

Detailed specifications with respect to the analysis of data from dual task studies are beyond the scope of this report. The reader is advised to consult appropriate sources on multivariate statistics (e.g., Pedhazur, 1982) and dual task methodology (e.g., Vidulich and Wickens, 1981; Wickens and Sandry, 1982).

TRAINING REQUIREMENTS

The subjects are presented with dual task instructions for the Memory Search and Unstable Tracking tasks. They are then told, as they will be performing both tasks at the same time, to remember that both tasks are equally important. Therefore, the object is to respond as quickly and accurately as possible on the Memory Search task while tracking as well as possible.

The first step of the training process requires that the tracking task and the Memory Search task each be performed alone until performance has

reached asymptote. Following this, dual task training can be started. Initial dual task performance is normally erratic. Thus, subjects should practice this task combination for a minimum of 15 minutes before any data are collected for analysis.

To summarize, the training phase for this test should consist of the following steps:

1. Read instructions to the subjects.

2. Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3. Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS (MEMORY SEARCH TASK)

The memory search task consists of two parts. In the first part of the task, you will be memorizing a small set of letters from the alphabet. This is called the "memory set." In the second part of the task, you will see a series of letters presented one at a time. Your task is to decide whether each letter is one of the letters in the memory set. If a letter is one of the memory set items, you press the "yes" key with your right hand; if it is not one of the memory set items, you press the "no" key with your right hand. The object of the task is to respond to the letters as quickly as possible without making any errors. Respond as fast as you can to the letters, but if you find yourself making errors, slow down. You should try to respond correctly to every item.

There will be either one, two, four, or six letters in the memory set. On some trials, you will have as much time as you need to memorize the letters

in the memory set. On other trials, this time will be set for you. It should take you not more than 15 to 20 seconds to commit the items to memory. The actual letters in the memory set will be different on each trial, so you will have to memorize a new set at the beginning of each trial. On certain trials only one probe letter will follow the memory set, on other trials 10 probes or 100 probes will follow the memory set.

INSTRUCTIONS TO SUBJECTS (UNSTABLE TRACKING TASK)

The object of the unstable tracking task is to keep a cursor centered over a target area in the middle of the screen of a CRT. You can control the movement of the cursor by turning the control knob with your left hand. Rotating the knob to the right (clockwise) moves the cursor up, and rotating it to the left (counterclockwise) moves it down. The cursor appears at the center of the screen and naturally tends to move vertically away from the center. Try to keep the cursor centered over the target at all times. If the cursor reaches the edge of the screen, it will reappear at the target and begin moving away again. This is called a control loss, and should be avoided if possible.

The task is run in 3-minute periods of data collection called trials. The difficulty of the control task will vary from trial to trial. During some trials, the cursor will be fairly easily kept in the middle of the screen, but others will be more unstable. To start the task, rotate the control knob until the numerical display on the screen reaches zero. The task automatically shuts off after 3 minutes and the screen will go blank.

MATCHING TO SAMPLE (UTC-PAB TEST NO. 24)
(SPATIAL MEMORY PATTERN RECOGNITION)

PURPOSE

This task is designed to assess the subject's ability to quickly and accurately choose a test stimulus which is identical to a standard stimulus presented previously. The test taps short term spatial memory and pattern recognition skills.

DESCRIPTION

The subject will be shown a single 4 by 4 matrix centered on the screen. The matrix will have cells of two colors (red and yellow). The number of cells of each color will be randomly determined for each stimulus. After viewing the sample stimulus for a time adequate for committing the stimulus to memory, the subject will initiate the presentation of the test trial. The test trial will consist of two 4 by 4 matrices, side by side on the screen. One of the matrices will be identical with the previously presented standard stimulus, while the other will be different. The subject's task is to select the test stimulus which matches the standard. There will be 30 such trials.

BACKGROUND

The matching to sample paradigm, first implemented in its present form by Skinner (1950), is designed to require the subject to maintain a standard in memory for some period of time (in this case, 1.5 seconds) before being offered a set of test stimuli for comparison (one of which matches the standard). After being offered the test stimuli, the subject is required to quickly and accurately decide which of them is identical to the standard. As a general rule, response times are on the order of 1000 msec. This task involves skills which fall into the realm of spatial ability.

The various facets of spatial ability can be arranged in a hierarchy (Lohman, 1979). One of the most useful graphic representations was presented in Figure 10. The factors can be characterized along two dimensions: speed/power and simplicity/complexity. The more powerful an ability, tne higher its position in the factor hierarchy. However, a higher position in the hierarchy also guarantees slower performance, since the tasks are more complex. At the top of the hierarchy is a factor called Visualization (Vz). It can best be thought of as the mental manipulation of a complex form or object in space. A second factor, found somewhat lower in the hierarchy, is called spatial orientation (SO). It is characteristic of tasks requiring the subject to imagine an object from a different vantage point. The third primary spatial factor (located still lower in the hierarchy) is called spatial relations (SR), and represents the ability to solve spatial problems quickly, by whatever means. There are four other spatial factors at the bottom of the hierarchy which deserve mention: Closure speed (Cs), the speed of matching incomplete or distorted stimuli with representations in long term memory; Kinesthetic (K), the speed of making left/right decisions; Visual memory (M), the ability to maintain stimuli in short term memory; and Perceptual speed (Ps), the speed of matching stimuli. The reader will note that all of these factors might play a part in the test under consideration here, with the possible exception of the kinesthetic factor. Thus, it is likely that this test will yield very quick reaction times, given that the factor loading appears to be concentrated on factors located low in the hierarchy.

Currently, one of the major problems in spatial perception research is the fact that little control is exercised over the subjects' choice of problem-solving strategies. With a small number of subjects, it is not difficult to evaluate each response to insure that the desired strategy is being used (i.e., for a Vz task, reorienting the imaginary object rather than the self). However, this problem becomes much greater as the number of subjects increases. With tests such as those in the UTC-PAB, it is safe to assume that the tests will be administered to large numbers of subjects; thus, it is important to consider the disparities induced in the data by the use of different strategies. Research has shown that more often than not, subjects use different strategies to solve the same test. Within a

test, the number of distinct strategies will increase as item difficulty and complexity increase. There will be a concomitant decrease in response speed as complexity increases. However, even on the most simple speeded tests, subjects still can be relied upon to use different strategies. Tests which the researcher intends to be solved using one strategy are often solved using another. For example, early researchers had great difficulty separating Vz and SO tests. It wasn't until they realized that SO tests were often solved using Vz strategies that the differentiation became more reliable. And finally, mental manipulation is often discarded in favor of more analytic methods as complexity and difficulty increase (i.e., the subjects may count angles or note distinctive features instead of using mental transformation to solve the problem).

It is obvious that various spatial abilities are present and available to the subject. However, caution must be used in any test of spatial ability. Tests are solved in different ways by different subjects. Instructions are only partially successful in guiding the subjects to use a specific strategy. Their solution strategies change as a function of various factors, including practice and item difficulty. Moreover, most factors represent individual differences in speed of solving particular types of problems, not general problem solving skills or abilities. Finally, the process of adapting a test to an experimental task may drastically alter the nature of the test. An experimental task will rarely tap exactly the same mental processes as the source test.

The current test involves 4 by 4 matrices made up of cells of two different colors. One of the most likely occurrences for this type of stimulus is that the subject will treat each pattern not as a two color figure, but as a brighter colored figure on a darker colored background or vice versa. The problem is, in effect, one of figure/ground in the classical Gestalt sense. Because of the nature of the problem, it may be appropriate to compare this problem to the various types of research done with dot patterns.

This UTC-PAB test involves same/different judgements based on the simultaneous presentation of two test patterns after the presentation of a

standard. The patterns (when evaluated from the viewpoint of a figure/ ground standpoint) are similar to those used by other researchers, including Ichikawa (1981), Klein and Armitage (1979), and Phillips (1974). The differences are worth noting, however. Ichikawa was studying ease of dot pattern memorization. He used 8-dot patterns in a 4 by 4 matrix, and 7-dot patterns in a 3 by 5 matrix. Through the use of a complicated metric, various types and levels of symmetry for each dot pattern were computed. These values were then applied (through multiple regression) to the results of a subjective rating of each pattern on a 9-point ease of memorization scale. The results were unequivocal: patterns which were rated as easy to memorize had much higher levels of symmetry than patterns which were rated as difficult to memorize. Implications for this study include possible differential responses based on the perceived symmetry of the standard and test patterns. Thus, it may be desirable to at least attempt to control for some of the more common types of symmetry.

Klein and Armitage (1979) used 7-dot patterns in a simultaneous pattern comparison task. It is unclear in what size matrix the dot pattern was embedded. Their study was intended to evaluate performance differences as a function of biological rhythms. These rhythms involved an alternation in the relative efficiency or activation of the two cerebral hemispheres. Klein and Armitage reasoned that, since the two hemispheres show differential specialization (e.g., spatial or verbal processing) frequent administration of two tests targeted for each hemisphere should demonstrate cyclical changes in performance. Their study showed just such a cycle, on the order of 90 minutes in length.

Phillips (1974) evaluated sensory storage and short term visual memory. His study is perhaps the most directly applicable to the current evaluation. He used matrices of three different sizes, four, six, or eight cells on a side. The density of dots was higher than in the other studies mentioned; the probability of a cell being filled was 0.5. He found that the 4 by 4 matrices had fairly long viable storage times (at least 9 seconds), losing no efficiency over the first 600 msec. In addition, the patterns tended to be quite resistant to masking or deficits induced by moving or shifting the pattern. In contrast, the larger matrices seemed to be stored

291

in the sensory store, and were markedly affected by movement, masking, and storage time. Storage time seemed to be limited to about 100 msec for the larger matrices. Thus, it appears that the choice of a 4 by 4 grid for the current study is the most viable one, based on the paradigm of choice.

Bridgeman and Mayer (1983) found that performance was at a chance level when subjects were required to shift fixation from one dot pattern position to another when trying to locate a single missing dot. The missing dot paradigm is similar to the current study's changing dot paradigm. Their patterns consisted of 12 dots in a 5 by 5 matrix that were presented under two separations (4 and 2.25 degrees). Implications for this UTC-PAB task suggest that presentation of the test stimuli as close as possible to the screen position of the standard may be the optimal presentation methodology.

RELIABILITY

Kennedy et al. (1985) quote the reliability of the Klein and Armitage (1979) task as 0.93 in their evaluation of several tests for inclusion in a portable microcomputer repeated measures testing system. In the Klein and Armitage task, the standard and test stimulus are presented simultaneously rather than successively as in the current experimental test. This makes it more difficult to generalize from that task to the current one, but little data is available otherwise.

VALIDITY

Again, the most similar test having computed validity data is the Klein and Armitage task. Research by Kennedy et al. (1985) has evaluated subjects' performance on this task in comparison with standardized tests of intelligence. The Klein and Armitage task correlated 0.57 with the WAIS performance scale, while correlating on 0.05 with the verbal scale. This implies that the task is not a verbal one. Within the subtests on the performance scale, the task correlates well with the spatial tests. The high correlations shown between the Klein and Armitage task suggest that it, too, is a spatial task.

SENSITIVITY

There is little data available on the effects of drugs, toxic agents, or environmental stressors on the specific test addressed in this manual. Other spatial tasks have been used in such studies, however, and may provide some indication of the possible effects of those factors on the current experimental task. The Manikin Test (which loads on the SO factor) (Carter and Woldstad, 1985) shows a severe performance decrement when administered to divers at extreme depth (Lewis and Baddeley, 1981; Logie and Baddeley, 1983). It is safe to assume that the Manikin Test also loads on other spatial factors, so it may be conjectured that a similar deficit would also occur with the present dot pattern presentation task.

TECHNICAL DESCRIPTION

The sample stimulus will be a square approximately 3.5 cm wide, centered on the screen. The stimulus will be subdivided into sixteen cells in a 4 by 4 matrix. The stimulus will be surrounded by a thin white border. In addition, this thin white border will also be present between the component cells of the stimulus. The color of each of the 16 cells in the sample stimulus is determined randomly, with the constraint that the ratio between the two colors is 7:9, 8:8, or 9:7. The limitation on the possible ratios helps to prevent the subject from matching-to-sample simply on the basis of color density for a given stimulus.

The sample stimulus is presented on the screen, and remains there until the subject presses any switch on the response box. The screen clears for 1.5 seconds and the two comparison stimuli are then presented. One of the test stimuli is identical to the standard, while the other has a single cell which is different. The difference is always in the location of the cell, not its color. Thus, if the lower right cell of the standard is red, the different matrix might have the position of that cell and a yellow cell elsewhere in the matrix swapped. In no case would the number of yellow cells be incremented. The process of swapping rather than replacing insures that the color ratios of the two stimuli remain the same.

The two comparison stimuli are presented with 3.5 cm between them, exactly the space occupied by the standard stimulus. On half of the 30 trials, the correct test stimulus will be on the left side of the screen, and on half the right. The position of the correct stimulus will be random across all subjects. The subject presses the corresponding button on the response box, following the subject's response the screen is cleared for 1 second, and the standard stimulus for the next trial is presented.

A single trial consists of the presentation of the standard stimulus, initiation of the test trial, presentation of the test stimulus pair, and an experimental response. If the initiation of the test stimulus pair does not occur within 60 seconds of the presentation of the standard stimulus, the test presentation will be initiated automatically. If the test presentation is not terminated by an experimental response within 60 seconds, the trial is terminated automatically, and the next trial begins.

Trial Specifications

Each trial will consist of the following sequence of events: (a) the standard stimulus will be presented for up to 60 seconds; (b) the screen will clear for 1.5 seconds; (c) the test stimulus pair will be presented for up to 60 seconds; (d) the subject will make a response; (e) during the training phase only, feedback on trial performance will be presented; and (f) the screen will clear and the next trial will be initiated.

DATA SPECIFICATIONS

Two separate response latency measurements will be recorded for each trial. The first will measure the time from the onset of the standard stimulus until the subject initiates the test presentation. The second measurement will record elapsed time from the onset of the test stimulus presentation until the subject makes his experimental response. These response latencies will be measured in milliseconds. The subject's response (either right or left) and the correct answer will also be recorded for each trial.

The following summary statistics will be computed after the session is complete: (a) percent correct responses; (b) the mean and median response latencies for the standard and test stimulus presentations; and (c) the range and variability of the standard and test stimulus presentations. It will be possible to examine the subject's data in a trial-by-trial format which will include the subject's response, the response latencies, and the correct response. It will be possible to examine all of the summary data on screen or via the printer.

TRAINING REQUIREMENTS

Initially, subjects should be read the instructions. After the instructions, the subjects should receive at least 10 trials of practice at the task to become familiar with it. During the training periods, there will be feedback after each trial. In other respects, the training trials will be identical to the experimental trials.

Since the instructions for this task stress fast and accurate performance, it is up to the experimenter to insure that the subject is optimizing his performance, (e.g., not sacrificing speed for accuracy or vice versa). If the experimenter feels that the subject does not understand the task or is performing incorrectly, additional instruction and test trials may be administered.

To summarize, the training phase for this test should consist of the following steps:

1.   Read instructions to the subjects.

2.   Run practice trials and evaluate subjects' performance to ensure that the instructions are being followed.

3.   Repeat the practice trials if it appears that the subjects require additional practice with the test.

4. Run the experimental trials. Note, if the tasks are being run over several sessions on this test, one may omit the practice trials after the first session.

INSTRUCTIONS TO SUBJECTS

During the course of this experiment, you will see a single matrix filled with red and yellow squares, followed by a pair of matrices. Your task is to decide which of the matrices in the pair match the single matrix you were shown first.

At the start of a trial you will see a single matrix made up of red and yellow cells. This is the sample matrix for the trial. You should do your best to memorize the pattern of red and yellow squares in this matrix. After you have memorized the sample matrix, press either button on the response box, and the sample matrix will be removed from the screen. After a short pause, you will then see two comparison matrices on the screen, side by side. One of these two matrices will be identical to the sample matrix that was on the screen, and the other matrix will differ slightly. Your task is to determine which of the two comparison matrices is the one which matches the sample matrix. If you think the matrix on the left matches the sample matrix, press the left button on your response box; if you think the matrix on the right matches the sample matrix, press the one on the right. You should try to decide which matrix matches the sample one as quickly as you can while still being accurate. If you have any questions, please ask the experimenter now.

Section 26
ITEM-ORDER TEST (UTC-PAB TEST NO. 25)
(SHORT TERM MEMORY RECOGNITION)

PURPOSE

The purpose of the item-order test is to examine a subject's ability to recognize strings of letters as being the same or different. Error rates produced from this test should reflect processes of short term memory recognition.

DESCRIPTION

In the item-order test, the subject sees a string of 7 consonants presented on the CRT. This is the target string. The target string is displayed for 2 seconds and then the CRT goes blank for 2.5 seconds. Immediately following the blank display, a new string of letters is presented. The second letter string is the test string. The subject is required to indicate whether the test string is identical to the target string. The subjects make their response by pressing one of two buttons. One button is labeled "same" and the other button is labeled "different." The test string bears one of three possible relationships to the target string: (1) the two strings are identical, (2) the same letters are in the two strings but the letters are in a different order, or (3) the two strings have different letters. Both of the previous cases qualify as "different." A single target string-test string pair constitutes one trial. The test consists of 40 trials. The dependent variables are response accuracy and response latency for each trial.

BACKGROUND

Recognition memory tasks, tasks involving judgements of identity and familiarity, are among the most common information processing tasks performed in everyday life (e.g., selecting the house key from one's keyring). Recognition memory can be described as the mental comparison of a present stimulus (the test stimulus) with the memorial representation of another (the target

stimulus). Mental comparisons may be either of two basic types. In the first, the respondent is asked to simply name the test stimulus, usually under impoverished viewing conditions. In the second type of recognition, the respondent is asked whether the test stimulus is familiar (i.e., has the test string been seen or heard before). This type of memory recognition is commonly examined using some variant of a string matching task.

Current theory regarding recognition memory dictates that test stimuli are presumed to be evaluated by the human respondent in terms of the familiarity attribute (knowledge of prior occurrence). It is commonly accepted that the level of this attribute, relative to some criterion value, determines whether a test stimulus is regarded by the respondent as familiar or not. One theory suggests that familiarity is a function of the frequency with which a stimulus has been perceived: Recognition judgements are based on the judged frequency of prior occurrence of the target stimulus (Underwood, 1983). Others have proposed that familiarity is mediated by intraitem organization, sensory, and perceptual integrations of the elements of the target stimulus (Mandler, 1980). It follows that any changes in the perceptual aspects of a stimulus should alter familiarity and recognition accuracy. The string matching paradigm allows control over these variables, enabling the researcher to determine what specific attributes of the target and test stimuli are encoded and retained in order to permit one stimulus to be distinguished from another.

In string matching, the subject hears or sees two series of items in immediate succession and is asked to decide whether the two series were or were not identical. To be judged identical the two strings must contain exactly the same items in exactly the same order. To be different, the strings might consist of one or more different items, or items might occur in different orders, or both of these two conditions. The UTC-PAB item-order test is a particular string matching task. Although no data has been published on this version of the test, experiments have been published using string matching tasks similar to the item-order.

Jahnke (in press) conducted several experiments using a string matching task. In the first experiment, if target and test strings differed, it was

298

only that one of the strings involved a transposition of two of the letters. The location of transposed letters vaired systematically. A total of 160 pairs of 7-letter strings were presented at a 2-letter per second rate with a 2-second interval between letter strings and a 5-second inter-trial silent interval during which subjects recorded their responses. Strings were composed of letters chosen to be phonologically dissimilar. It was expected that error rates would vary according to the location of the transposed letters, since there is evidence that the phonological properties of the target letters and the locations of the letters in the string are important memory attributes (Drewnowski, 1980).

The results for the pairs with transposed letters indicate that error rates are highest when certain adjacent letters are transposed. In the lay zero conditions (zero letters separate the transposed letters), performance was poorest for the transposition either earliest (condition two and three, 27 percent errors) or latest (condition five and six, 29 percent errors) in the string. Performance on letters at the same lag in the middle of the string was relatively good. Also, performance was good for strings in which letters in position five or six were transposed with a letter most distant from it (high lag value). Thus, it can be concluded that serial position and lag play an important role in recognition memory.

The second experiment was designed to determine how sensitive respondents are to test strings that differ from the target by the substitution of one or more new letters (e.g., FHJXLNQ-FHJRLNQ). Because one or more new phonologically distinct letters are introduced in the test string, the respondent should often correctly identify "different" pairs as "different" when the stimuli are presented auditorially. However, recognition errors are expected and the error rates should vary according to the location of the substituted item(s). The results for strings that differed by a single letter had an average error rate of 17 percent over the five possible serial positions. Statistical analyses showed that none of the serial position entries differed significantly from any other. Thus, in this experiment, serial position of a substituted letter was not an effective variable. When more than one letter is substituted, the error rates become lower. Thus, the analysis of error rates in a string matching task assists

299

in the understanding of basic recognition process which are critically involved in all sorts of natural situations, including the recognition of faces, listening, and reading.

Another study conducted by Eichelman (1970) compared recognition performance of words to that of letter strings. Recognition of words and letter strings of the same lengths (either 1, 2, 4, or 6 letters) were performed in order to determine the effect of a familiarity (words) attribute on recognition memory. Results showed that the number of letters had a significant effect on the number of errors where the obtained error rates for 1-, 2-, 4-, and 6-letter strings were 5 percent, 4.6 percent, 9.1 percent, and 6.1 percent, respectively. Also, word strings were matched significantly faster than letter strings for four and six letters. Thus, the familiarity of words significantly increased reaction time but did not have an effect on the number of errors. The number of errors was significantly affected by the number of letters only and not familiarity.

RELIABILITY

It is important for any test to possess a degree of consistency or stability of scores across trials and sessions. This consistency is known as test-retest reliability and is a measure of the degree to which performance on the test remains constant over different testing sessions. Unfortunately, no reliability studies have been conducted for string matching tasks thus far. Therefore, there is no indication of how results obtained on one session of the item-order test will resemble the results of other sessions. This information would also reflect the point at which performance stabilizes and further practice has no effect on performance. A study involving performance of the item-order test for a number of sessions for 15 consecutive days would provide the necessary test-retest reliability information for this test.

VALIDITY

The item-order test is designed to place variable demands on short term recognition memory. By replacing an item and varying the order of an item

in a list, different recognition errors may occur reflecting these different memory processes. Significantly different recognition errors have been reported as a function of serial position in the list for a transposed item and also for a replaced item (Jahnke, in press). Recognition memory has also been shown to be dependent on familiarity of the strings and the number of items making up a string (Eichelmen, 1970). Although the procedure is very similar, no data has been collected on the item-order test to determine if recognition memory processes are affected. Thus, the validity of the item-order test as a test of memory recognition must remain uncertain until data can be collected and discussed in relation to findings of similar string matching tasks.

SENSITIVITY

Investigations involving the performance of string matching tasks under the presence of environmental stressors have not been reported in the literature to date. Research investigating the effects of sleep loss or drugs (e.g., diazepam, atropine, alcohol) on short term memory recognition via the item-order test, would be appropriate and useful. Research testing the effects of these variables on other short term memory processes (comparison, recall) has been reported in the literature (e.g., Smith and Langolf, 1981; see UTC-PAB Manual No. 9: Memory Search). Although the effects of drugs on these short term memory processes have been well documented, recognition processes may differ from recall processes and, thus, may be affected in a different manner.

TECHNICAL DESCRIPTION

The letters in both the target and test strings are one inch high and are in upper case format. The string is displayed centered on the CRT. The strings are restricted to consonants. The consonants for each target string are randomly selected from the pool of all English consonants. Each string is made up of seven letters. The test is composed so that half of the trials require a "same" response and half of the trials require a "different" response. The "different" trials are half item-different and half order-different. An item-different trial is one where the test string has

301

one new letter in it that replaced a letter that was in the target string.
An order-different trial is one where the test string has two items inter-
changed in their original position as compared to the original order in the
target string. In the order-different strings the letters that are inter-
changed are always contiguous. The letters that are replaced or inter-
changed are selected randomly for each trial, with the restriction that the
first and last letters in the target string are never changed in the test
string. The occurrence of the "same" and "different" trials in the test is
determined randomly.

Trial Specifications

The test consists of 40 trials (20 "same" and 20 "different" trials). A
trial consists of the presentation of one target string and its corres-
ponding test string. The target string is presented for 2 seconds. The
CRT is blanked for 2.5 seconds followed by the presentation of the test
string. Following the subject's response to a test string, a row of stars
is displayed for 500 msec to signal the start of a new trial.

DATA SPECIFICATIONS

The subject's response accuracy and response latency for each trial will be
recorded. The measurement of response latency begins with the presentation
of the test string and concludes when the subject presses a response but-
ton. Response latency is measured with an accuracy of 1 msec. Response
accuracy is simply whether the response is correct. Completed summary
statistics include the total number of correct responses made on the test,
the number of correct responses made on the "same" trials, the number of
correct responses made on the "item-different" trials, and the number of
correct responses made on the "order-different" trials. The median and
mean response latency for the entire test is provided as well as the median
and mean response latency for the "same" trials, "item-different" trials,
and "order-different" trials.

302

TRAINING REQUIREMENTS

The instructions should be read to the subjects at the beginning of each
testing session. The training for this test is as follows: 10 practice
trials are given to the subjects following the same procedures as used in
the test proper. However, when a subject makes an incorrect response to
one of the training trials, the message "That was incorrect" will appear.
The target string and the test string will be displayed directly below the
message. This feedback screen will be presented for 5 seconds and then the
next practice trial will commence.

To summarize, the training phase for this test should consist of the
following steps:

1.   Read instructions to the subjects.

2.   Run practice trials and evaluate subjects' performance to ensure that
     the instructions are being followed.

3.   Repeat the practice trials if it appears that the subjects require
     additional practice with the test.

4.   Run the experimental trials. Note, if the tasks are being run over
     several sessions on this test, one may omit the practice trials after
     the first session.

INSTRUCTIONS TO SUBJECTS

You will see displayed on the computer screen a string of seven letters for
a short time (2 seconds). Study the letters quickly so that you will
remember what letters were on the screen and the order in which they
appeared. The screen will go blank for a short time and then you will see
seven more letters. Your task is to decide whether these seven letters are
exactly the same as the seven letters you just studied. If the two strings
are identical, press the button labeled "same." However, if either (1)
there is a letter in the test string that wasn't in the original string you

studied, or (2) the letters are in a different order than they were when you studied them, indicate this difference by pressing the button labeled "different." In any case, please press a button as quickly as possible without making errors. After you have pressed a button, some stars will appear briefly on the screen; these stars mean that you should prepare to study a new string of letters which will soon appear.

APPENDIX A

Appendix A presents modifications to the UTC-PAB Tests that are presented
in the proposed UTC-PAB: Review and Methodology.

UTC-PAB Test No. 6

The new version of the Continuous Recognition Test contains the following
modification relative to the version proposed for the UTC-PAB:

- The three difficulty levels are defined by the number of posi-
  tions that must be maintained in memory--1, 2, or 3 positions
  back. In the new version the subjects will only match single
  digit numbers.

The above modification is based on the results of recent research conducted
at AAMRL (the results of this research have not been published). The study
included 12 subjects that were tested on four consecutive days. On each
day the subjects performed four 3-minute trials for each difficulty condi-
tion (1, 2, or 3 positions back). The following summary statistics are the
average number of correct digit recognitions for the fourth day of testing:

Positions Back	Average Percent Correct
1	96.73
2	94.13
3	89.69

The average percent correct for digit recognitions decreased as a function
of the number of positions back. The differences among the three condi-
tions are statistically significant. The recommended performance metric
for this task is the percent of correct digit recognitions per 3-minute
trials.

The new version of this test presents a significant improvement relative to
the version that was originally recommended for inclusion in the UTC-PAB.
The new version presents three levels of difficulty that are generated

through the manipulation of a single variable (e.g., number of positions back). Whereas the original version of the test involves the manipulation of two different variables (number of digits and number of positions back) in an unsystematic fashion. Since these two variables are not manipulated systematically, it is not possible to unambiguously determine which of the two variables (number of digits or number of positions back) is causally related to a given performance decrement due to the effect of treatment or pretreatment drug.

UTC-PAB Test No. 17

The new version of the Visual Probability Monitoring Test contains the following modificaticns relative to the version that was originally proposed for inclusion in the UTC-PAB:

- Pointer update rate was increased from 2 per second to 5 per second.

- The number of signals was increased to 10 per 3-minute trial.

- The difficulty levels are defined by the number of dials: 1, 2, or 3 dials. The bias for signal pointer moves is 95 percent for all three conditions.

The above modifications are based on current research conducted at AAMRL (Eggemeier and Ammel, 1986). The study included 12 subjects that were tested on four consecutive days. On each day the subjects performed four 3-minute trials for each difficulty condition (1, 2, or 3 dials). The following summary statistics are the average reaction times for detecting signals (e.g., biased pointer movements):

Number of Dials	Average Reaction Time (Seconds)
1	3.54
2	4.33
3	5.07

The above difficulty levels represent conditions that are statistically different with respect to reaction times to the detection of signals. Also, the error rates for the three conditions were relatively low (less than 10 percent). The recommended performance metric for the new version of the Visual Probability Monitoring Test is the reaction time to signal detection.

The new version of this test presents significant improvements relative to the version that was originally recommended for inclusion in the UTC-PAB. The improvements are as follows: (a) the increase in the number of signals per trial allows the use of parametric statistical tools for the evaluation of performance (the original version resulted in only three or less signals per trial and the performance measures did not meet the requirements for parametric analysis); and (b) the manipulation of task difficulty is accomplished by only varying the number of signals (1, 2, or 3 dials) rather than varying number of signal sources and signal bias. The manipulation of number of dials and signal bias simultaneously presented difficulties with respect to the interpretation of performance decrements in this task. Since these two variables are not manipulated in a systematic fashion, it is not possible to unambiguously determine which of the two variables (number of dials or signal bias) is causally related to a given performance decrement due to the effect of treatment or pretreatment drug. The new version of this test does not present the above interpretation problem since only one variable (number of dials) is systematically manipulated to produce the three difficulty levels.

UTC-PAB Test No. 22

The new version of the Unstable Tracking Test contains the following modifications relative to the version originally proposed for inclusion in the UTC-PAB:

- The difficulty levels are lambdas of 1, 2, and 3 for the low, medium, and high difficulty conditions.

• The tracking cursor moves in a horizontal direction rather than vertically as in the original version of the test.

The above modifications are based on the results of recent research conducted at AAMRL (the results of this research have not been published). The study included 12 subjects that were tested on four consecutive days. The subjects performed four 3-minute trials for each difficulty condition (lambdas of 1, 2, or 3). The following summary statistics are the average number of edge violations and RMS error for the fourth day of testing:

Lambda	Average Number of Edge Violations	Average RMS Error
1	0.26	7.09
2	9.29	22.06
3	48.75	34.98

The average number of edge violations and RMS error increased as a function of the value of lambda. RMS error is the recommended metric for this test. The differences between the three difficulty conditions are statistically reliable and the relationship between RMS error and lambda is linear.

The new version of the Unstable Tracking Test presents an improvement relative to the version that was originally proposed. The new version presents the tracking stimulus such that operator inputs and stimulus movements are mapped in a compatible manner (e.g., a leftward movement of the tracking controller translates to a leftward movement of the tracking cursor). Also, the difficulty levels represent increments in task demand that are evenly spaced (the original version used lambda values of 1, 2, and 5). Also, the above improvements present three levels of tracking difficulty that require nearly the same amount of training to reach stable performance (twelve 3-minute trials per condition).

REFERENCES

Adam, N. and G. Collins, 1978, Late Components of the Visual Evoked Potential to Search in Short Term Memory, Electroencephalography and Clinical Neurophysiology, 44, 147-156.

Adler, A., H. Strasser, and W. Muller-Limmroth, 1976, Effects of Massed and Distributed Practice, Additional Monetary and Placebo Incentives, and Changes in Characteristics of Control Element During Learning of Human Motor Control, Perceptual and Motor Skills, 42, 1335-1338.

Aitken, R. C. B. and J. L. Gedye, 1968, A Study of Two Factors Which Affect Arousal Level and the Apparent Duration of a Ten-minute Interval, British Journal of Psychology, 59, 253-263.

Alexander, D. A., 1973, Attention Dysfunction in Senile Dementia, Psychological Reports, 32, 229-230.

Anders, T. R., J. L. Fozard, and T. D. Lillyquist, 1972, Effects of Age Upon Retrieval from Short Term Memory, Developmental Psychology, 6, 214-217.

Angus, R. G. and R. J. Heslegrave, 1985, Effects of Sleep Loss on Sustained Cognitive Performance during a Command and Control Simulation, Behavior Research Methods, Instruments and Computers, 17, 55-67.

Ashcraft, M. H. and J. Battaglia, 1978, Cognitive Arithmetic: Evidence for Retrieval and Decision Processes in Mental Addition, Journal of Experimental Psychology: Human Learning and Memory, 4, 527-538.

Ashcraft, M. H. and E. H. Stazyk, 1981, Mental Addition: A Test of Three Verification Models, Memory and Cognition, 9, 185-196.

Ashkenas, I. L. and D. T. McRuer, 1959, The Determination of Lateral Handling Quality Requirements from Airframe/Human-pilot System Studies, Wright Air Development Center, Wright-Patterson AFB, Ohio, TR 59-135.

Atkinson, A. C. and D. Whitfield, 1972, Hovercraft Control Skills, Occupational Psychology, 46, 79-86.

Atkinson, R. C., S. E. Holmgren, and J. F. Juolea, 1969, Processing Time as Influenced by the Number of Elements in a Visual Display, Perception and Psychophysics, 6, 321-327.

Attneave, F., 1955, Symmetry, Information, and Memory for Patterns, American Journal of Psychology, 68, 209-222.

Baddeley, A. D., 1968, A 3-Minute Reasoning Test Based on Grammatical Transformations, Psychonomic Science, 10, 341-342.

Baddeley, A. D., W. J. Cuccaro, G. H. Egstrom, G. Weltman, and M. A. Willis, 1975, Cognitive Efficiency of Divers Working in Cold Water, Human Factors, 17, 446-454.

Baddeley, A. D. and N. C. Flemming, 1967, The Efficiency of Divers Breathing Oxyhelium, Ergonomics, 10, 311-319.

Baddeley, A. D. and G. J. Hitch, 1974, Working Memory, In G. H. Bower (Ed.), The Psychology of Learning and Motivation (Vol. 8), New York: Academic Presses.

Benson, A. J. and J. L. Gedye, 1963, Logical Processes in the Resolution of Orientation Conflict (Report 259), Farnborough, United Kingdom: Royal Air Force, Institute of Aviation Medicine.

Bonnet, M. H., 1980, Sleep, Performance and Mood After the Energy-Expenditure Equivalent of 40 Hours of Sleep Deprivation, Psychophysiology, 17, 56-63.

Bridgemen, B. and M. Mayer, 1983, Failure to Iterate Visual Information from Successive Fixations, Bulletin of the Psychonomic Society, 21, 285-286.

Briggs, G. E. and J. Blaha, 1969, Memory Retrieval and Central Comparison Times in Information Processing, Journal of Experimental Psychology, 79, 395-402.

Briggs, G. E. and A. M. Johnsen, 1973, On the Nature of Central Processing in Choice Reactions, Memory and Cognition, 1, 91-100.

Briggs, G. E., G. L. Peters, and R. P. Fisher, 1972, On the Locus of the Divided-attention Affects, Perception and Psychophysics, 11, 315-320.

Briggs, G. E. and J. M. Swanson, 1969, Retrieval Time as a Function of Memory Ensemble Size, Quarterly Journal of Experimental Psychology, 21, 185-191.

Broadbent, D. E., 1958, Perception and Communications, London: Pergamon Press.

Brookhuis, K. A., G. Mulder, L. J. M. Mulder, A. B. M. Gloerich, H. J. Van Dellen, J. J. Van Der Meere, and H. Ellermann, 1981, Late Positive Components and Stimulus Evaluation Time, Biological Psychology, 13, 107-123.

Brown, I. D., D. C. V. Simmonds, and A. H. Tickner, 1967, Measurement of Control Skills, Vigilance, and Performance on a Subsidiary Task During 12 Hours of Car Driving, Ergonomics, 10, 665-673.

Brown, I. D., A. H. Tickner, and D. C. Simonds, 1968, Interference Between Concurrent Tasks of Driving and Telephoning, Journal of Applied Psychology, 53, 419-424.

Burrows, D. and R. Okada, 1971, Serial Position Effects in High-speed Memory Search, Perception and Psychophysics, 10, 305-308.

Burrows, D. and R. Okada, 1973, Parallel Scanning of Semantic and Formal Information, Journal of Experimental Psychology, 97, 254-257.

Bush, M., 1977, The Relationship Between Impaired Selective Attention and Severity of Psychopathology in Acute Psychiatric Patients, British Journal of Medical Psychology, 50, 251-265.

Callaway, E. and R. I. Band, 1958, Some Psychopharmacological Effects of Atropine, AMA Archives of Neurology and Psychiatry, 79, 91-102.

Canestrari, R. E., 1962, The Effects of Aging on Vigilance Performance, Paper Presented to a Meeting of the Gerontological Society, Miami, Florida.

Carpenter, A. and R. Conrad, 1953, An Apparatus Providing Signals at "Random" Intervals, Quarterly Journal of Experimental Psychology, 5, 87-88.

Carpenter, J. A. and B. M. Ross, 1965, Effect of Alcohol on Snort Term Memory, Quarterly Journal of Studies on Alcohol, 26, 561-579.

Carter, R. C. and R. J. Biersner, 1982, US Navy Enlisted Jobs: An Analysis, (Research Report No. NBDL-82R010), New Orleans, Louisiana: Naval Biodynamics Laboratory, (NTIS No. ADA 122410).

Carter, R. C., R. S. Kennedy, and A. C. Bittner, Jr., 1981, Grammatical Reasoning: A Stable Performance Yardstick, Human Factors, 23, 587-591.

Carter, R. C., R. S. Kennedy, A. C. Bittner, Jr., and M. Krause, 1980, Item Recognition as a Performance Evaluation Test for Environmental Research, Proceedings of the 24th Annual Meeting of the Human Factors Society, Santa Monica, California: Human Factors Society, 340-344.

Carter, R. C. and M. Krause, 1983, Reliability of Slope Scores for Individuals, Naval Biodynamics Laboratory, New Orleans, (Research Report No. NBDL-82R002, NTIS, No. AD A115030).

Carter, R. C. and M. Krasue, 1983, Reliability of Slope Scores for Individuals: Experiment 4: Letter Search, Naval Biodynamics Laboratory, New Orleans, (Research Report No. NBDL-83R003) 17-20.

Carter, R. and J. Woldstad, 1985, Repeated Measures of Spatial Ability with the Manikin Test, Human Factors, 27, 209-219.

Casali, J. G. and W. W. Wierwille, 1983, A Comparison of Rating Scale, Secondary Task, Physiological, and Primary Task Workload Estimation Techniques in a Simulated Flight Task Emphasizing Communications Load, Human Factors, 25, 623-641.

Casali, J. G. and W. W. Wierwille, 1984, On the Measurement of Pilot Perceptual Workload: A Comparison of Assessment Techniques Addressing Sensitivity and Intrusion Issues, Ergonomics, 27, 1033-1050.

Chase, W. G. and R. C. Calfe, 1969, Modality and Similarity Effects in Short Term Recognition Memory, Journal of Experimental Psychology, 81, 510-514.

Chase, W. G. and H. H. Clark, 1972, Mental Operations in the Comparison of Sentences and Pictures, In L. W. Gregg (Ed.), Cognition in Learning and Memory, New York: Wiley and Sons.

Checkosky, S. F. and D. Whitlock, 1973, The Effect of Pattern Goodness in the Reproduction of Backword-Masked Patterns, Journal of Experimental Psychology: Human Perception and Performance, 2, 139-150.

Cherry, C., 1953, Some Experiments on the Reception of Speech With One and With Two Ears, Journal of the Acoustical Society of America, 25, 975-979.

Cherry, N., J. D. Johnston, H. Venables, H. A. Waldron, L. Buck, and C. J. Mackay, 1983, The Effects of Toluene and Alcohol on Psychomotor Performance, Ergonomics, 26, 1081-1087.

Chiles, W. D., 1977, Objective Methods for Developing Indices of Pilot Workload, Report No. FAA-AM-77-15, Washington, D.C.: U.S. Department of Transporation, Federal Aviation Administration.

Chiles, W. D., E. A. Alluisi, 1979, On the Specification of Operator or Occupational Workload with Performance-Measurement Methods, Human Factors, 21, 515-528.

Chiles, W. D., E. A. Alluisi, and O. S. Adams, 1968, Work Schedules and Performance During Confinement, Human Factors, 10, 143-196.

Chiles, W. D., C. B. Bruni, and R. A. Lewis, 1969, Methodology in the Assessment of Complex Human Performance: The Effects of Dignal Rate on Monitoring a Dynamic Process, Report No. FAA-AM-69-9, Washington, D.C.: U.S. Department of Transportation, Federal Aviation Administration.

Chiles, W. D. and A. E. Jennings, 1970, Effects of Alcohol on Complex Performance, Human Factors, 12, 605-612.

Chiles, W. D., A. E. Jennings, and E. A. Alluisi, 1978, The Measurement and Scaling of Workload in Complex Performance, Report No. FAA-AM-7834, Washington, D.C.: U.S. Department of Transportation, Federal Aviation Administration.

Clark, H. H. and W. G. Chase, 1972, On the Process of Comparing Sentences Against Pictures, Cognitive Psychology, 3, 472-517.

Clark, H. H. and W. G. Chase, 1974, Perceptual Coding Strategies in the Formation and Verification of Descriptions, Memory and Cognition, 2, 101-111.

Clement, D. E., 1967, Paired Associated Learning as a Correlate of Pattern Goodness, Journal of Verbal Learning and Verbal Behavior, 2, 427-431.

Clifton, C., Jr. and J. Tash, 1973, Effect of Syllabic Word Length on Memory-Search Rate, Journal of Experimental Psychology, 99, 231-235.

Cohen, J., 1957a, A Factor-analytically Based Rationale for the Wechsler Adult Intelligence Scale, Journal of Consulting Psychology, 21, 451-457.

314

Cohen, J., 1957b, The Factorial Structure of the WAIS Between Early Adult-hood and Old Age, Journal of Consulting Psychology, 21, 283-290.

Cohen, V. V. R. and R. W. Pew, 1970, Small Size Targets Sets in Visual Search Under Acuracy Set, Paper Presented at the Meeting of the Midwestern Psychological Association, Cincinnati.

Collins, A. M. and E. F. Loftus, 1975, A Spreading-Activation Theory of Semantic Processing, Psychological Review, 82, 407-428.

Connor, J. M., 1972, Serial and Parallel Encoding Processes in Memory and Visual Search, Journal of Experimental Psychology, 96, 363-370.

Conrad, R., 1955, Some Effects on Performance of Changes in Perceptual Load, Journal of Experimental Psychology, 49, 313-322.

Cooper, L. A. and R. N. Shepard, 1978, Transformations on Representations of Objects in Space, In E. C. Carterette and M. P. Friedman (Eds.), Handbook of Perception, Volume 8: Space and Object Perception, New York: Academic Press.

Craik, F. I. M. and E. Tulving, 1975, Depth of Processing and the Retention of Words in Episodic Memory, Journal of Experimental Psychology: General, 104, 268-294.

Craik, F. I. M. and M. J. Watkins, 1973, The Role of Rehearsal in Short Term Memory, Journal of Verbal Learning and Verbal Behavior, 12, 599-607.

Crosby, J. V. and G. R. Parkinson, 1979, A Dual Task Investigation of Pilots' Skill Level, Ergonomics, 22, 1301-1313.

Crowell, E. B. and J. S. Ketchum, 1967, The Treatment of Scopolamine-Induced Delirium with Physostigmine, Clinical Pharmacology and Therapeutics, 8, 409-414.

Damos, D. L., A. C. Bittner, R. S. Kennedy, and M. M. Harbeson, 1981, Effects of Extended Practice on Dual Task Tracking Performance, Human Factors, 23, 627-631.

Damos, D. L., A. C. Bittner, R. S. Kennedy, M. M. Harbeson, and M. K. Krause, 1984, Performance Evaluation Tests for Environmental Research (PETER): Critical Tracking Test, Perceptual and Motor Skills, 58, 567-573.

Damos, D. L. and E. A. Lyall, 1984, The Effect of Asymmetric Transfer on Dual Task Assessment of Voice Technology, Proceedings of the Human Factors Society 28th Annual Meeting, Volume 1, 21-26.

Darley, C. F., R. L. Klatzky, and R. C. Atkinson, 1972, Effects of Memory Load on Reaction Time, Journal of Experimental Psychology, 96, 232-234.

Davies, A. D. M. and D. R. Davies, 1975, The Effects of Noise and Time of Day Upon Age Differences in Performance at Two Checking Tasks, Ergonomics, 18, 321-336.

Derner, G. F., M. Aborne, and A. M. Castore, 1950, Journal of Consulting Psychology, 14, 172-179.

DeRosa, D. V., 1969, Transformation on Sets in Short Term Memory: Set Size Reduction by Deletion, Journal of Experimental Psychology, 82, 415-426.

DeRosa, D. V. and R. E. Morin, 1970, Recognition Reaction Time for Digits in Consecutive and Nonconsecutive Sets, Journal of Experimental Psychology, 83, 472-479.

DeRosa, D. V. and M. Sabol, 1973, Transformations on Sets in Short Term Memory: Temporal and Spatial Factors Influencing Deletion, Memory and Cognition, 1, 69-72.

Donders, F. C., 1969, On the Speed of Mental Processes, (Translated by W. G. Koster), Acta Psychologica, 30, 412-431.

Dott, A. B. and R. K. McKelvy, 1977, Influence of Ethyl Alcohol in Moderate Levels on Visual Stimulus Tracking, Human Factors, 19, 191-199.

Drewnowski, A., 1980, Attributes and Priorities in Short Term Recall: A New Model of Memory Span, Journal of Experimental Psychology: General, 109, 208-250.

Duncan-Johnson, C. C. and B. S. Kopell, 1981, The Stroop Effect: Brain Potentials Localize the Source of Interference, Science, 214, 938-940.

Dyer, F. N., 1973, The Stroop Phenomenon and Its Use in the Study of Perceptual, Cognitive, and Response Processes, Memory and Cognition, 1, 106-120.

Egan, D. E., 1978, Characterizing Spatial Ability: Different Mental Processes Reflected in Accuracy and Latency Scores, (Research Report No. 1250), Pensacola, Florida: Naval Aerospace Medical Research Laboratory.

Eggemeier, F. T. and J. R. Amell, 1986, Visual Probability Monitoring: Effects of Display Load and Signal Discriminability, Proceedings of the Human Factors Society, Volume 1, 63.

Eichelmen, W. H., 1970, Familiarity Effects in the Simultaneous Matching Task, Journal of Experimental Psychology, 86, 275-282.

Englund, C. E., P. Naitoh, and D. Ryman, 1984, Effects of Moderate Physical Work on Chronopsychological Variables During Sustained Performance, In E. Hause and H. F. Kabat (Eds.), Chronobiology, Basel, Switzerland: Karger.

Englund, C. E., P. Naitoh, D. H. Ryman, and J. A. Hodgdon, 1983, Moderate Physical Work Effects on Performance and Mood During Sustained Operations (SUSOPS), (NHRC Tech Report No. TR-83-6), San Diego, California: NHRC.

Englund, C. E., R. L. Reeves, C. A. Shingledecker, D. R. Thorne, K. P. Wilson, and F. W. Hegge, 1985, Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB), I: Design and Specification of the Battery,

JW6D3 MILPERF Report No. 85-1, U.S. Army Research and Development Command, Fort Detrick, Maryland.

Englund, C. E., D. H. Ryman, P. Naitoh, and J. A. Hodgdon, 1985, Cognitive Performance During Successive Sustained Physical Work Episodes, Behavior Research Methods, Instruments, and Computers, 17, 75-85.

Eriksen, D. W. and W. P. O'Hara, 1982, Are Nominal Same/Different Matches Slower Due to Differences in Level of Processing or to Response Competition, Perception and Psychophysics, 32, 335-344.

Eriksen, C. W., W. P. O'Hara, and B. A. Eriksen, 1982, Response Competition Effects in Same/Different Judgments, Perception and Psychophysics, 32, 261-270.

Fitts, P. M., M. Weinstein, M. Rappaport, N. Anderson, and J. A. Leonard, 1956, Stimulus Correlates of Visual Pattern Perception: A Probability Approach, Journal of Experimental Psychology, 51, 1-11.

Flowers, J. H., 1975, "Sensory" Interference in a Word-Color Matching Task, Perception and Psychophysics, 18, 37-43.

Folkard, S., 1975, Diurnal Variation in Logical Reasoning, British Journal of Psychology, 66, 1-8.

Ford, J. M., W. T. Roth, R. C. Mohs, W. F. Hopkins, and B. S. Kopell, 1979, Event Related Potentials Recorded from Young and Old Adults During a Memory Retrieval Task, Electroencephalography and Clinical Neurophysiology, 47, 450-459.

Foss, D. J. and B. E. Dowell, 1971, High-Speed Memory Retrieval with Auditorily Presented Stimuli, Perception and Psychophysics, 9, 465-468.

French, J. W., R. B. Ekstrom, and L. A. Price, 1963, Manual for Kit of Reference Tests for Cognitive Factors, Educational Testing Services, Princeton, New Jersey.

Glanzer, M., T. Taub, and R. Murphy, 1968, An Evaluation of Three Theories of Figural Organization, _American Journal of Psychology_, 81, 53-66.

Glenville, M., R. Broughton, A. M. Wing, and R. T. Wilkinson, 1978, Effects of Sleep Deprivation of Short Duration Performance Measures Compared to the Wilkinson Auditory Vigilance Task, _Sleep_, 1, 169-176.

Glenville, M. and R. T. Wilkinson, 1979, Portable Devices for Measuring Performance in the Field: The Effects of Sleep Deprivation and Night Shift on the Performance of Computer Operators, _Ergonomics_, 22, 927-933.

Gomer, F. E., R. J. Spicuzza, and R. D. O'Donnel, 1976, Evoked Potential Correlates of Visual Item Recognition During Memory Scanning Tasks, _Physiological Psychology_, 4, 61-65.

Gopher, D., 1982, A Selective Attention Test as a Predictor of Success in Flight Training, _Human Factors_, 24, 173-183.

Gopher, D. and D. Kahneman, 1971, Individual Differences in Attention and the Prediction of Light Criteria, _Perceptual and Motor Skills_, 33, 1335-1342.

Gould, J. D. and R. Carn, 1973, Visual Search, Complex Backgrounds, Mental Counters, and Eye Movements, _Perception and Psychophysics_, 14, 125-132.

Graboi, D., 1971, Searching for Targets: The Effects of Specific Practice, _Perception and Psychophysics_, 10, 300-304.

Greenwald, A. G., 1979, Time-Sharing Ideomotor Compatability, and Automaticity, _Proceedings, 23rd Annual Meeting of the Human Factors Society_.

Grether, W. F., 1971, _Acceleration and Human Performance_, Technical Report No. AMRL-TR-71-22, Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Hall, T. J., G. E. Passey, and T. W. Meighan, 1965, _Performance of Vigilance and Monitoring Tasks as a Function of Workload_, Report No. AMRL-TR-65-22,

Wright-Patterson Air Force Base, Ohio, Aerospace Medical Research Laboratories.

Hann, R. L., 1973, The Sternberg Paradigm and Memory Scan Research, Major Area Paper, Vanderbilt University.

Harbeson, M. M., R. S. Kennedy, M. Krause, and A. C. Bittner, 1982, Repeated Measures of Information Processing, Proceedings of the 26th Annual Meeting of the Human Factors Society, Seattle, Washington.

Harbeson, M. M., M. Krause, R. S. Kennedy, and A. C. Bittner, 1982, The Stroop as a Performance Evaluation Test for Environmental Research, The Journal of Psychology, 111, 223-233.

Harris, G. J. and R. E. Fleer, 1974, High-Speed Memory Scanning in Mental Retardates: Evidence for a Central Processing Deficit, Journal of Experimental Child Psychology, 17, 452-459.

Haslam, D. R., 1985, Sleep Deprivation and Nap, Behavior Research Methods, Instruments and Computers, 17, 46-54.

Hegge, F. W., D. L. Reeves, D. P. Poole, and D. R. Thorne, 1985, The Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB): II: Hardware/Software Design and Specification, JW6D3 MILPERF Report No. 85-2, U.S. Army Research and Development Command, Fort Detrick, Maryland.

Hemsley, D. R. and P. M. Richardson, 1980, Shadowing by Context in Schizophrenia, The Journal of Nervous and Mental Disease, 168, 141-145.

Herbert, M., T. E. J. Healy, J. B. Bourke, I. R. Fletcher, J. M. Rose, 1983, Profile of Recovery After General Anaesthesia, British Medical Journal, 286, 1539-1542.

Herbert, M., P. J. Standen, A. H. Short, and A. T. Birmingham, 1983, A comparison of some Psychological and Physiological Effects Exerted by Zetidoline (DL 308) and by Oxazepam, Psychopharmacology, 81, 335-339.

Hick, W. E., 1952, On the Rate of Gain of Information, Quarterly Journal of Experimental Psychology, 87, 514-517.

Hicks, R. E. and G. W. Miller, 1976, Transfer Between Time Judgements of Various Durations as a Function of Information Feedback, American Journal of Psychology, 89, 303-310.

Hicks, R. E., G. W. Miller, and M. Kinsbourne, 1976, Prospective and Retrospective Judgements of Time as a Function of Amount of Information Processed, American Journal of Psychology, 89, 719-730.

Hitch, G. J., 1978, The Role of Short Term Working Memory in Mental Arithmetic, Cognitive Psychology, 10, 302-323.

Hitch, G. J. and A. D. Baddeley, 1976, Verbal Reasoning and Working Memory, Quarterly Journal of Experimental Psychology, 28, 603-621.

Hock, H. S. and H. Egeth, 1970, Verbal Interference with Encoding in a Perceptual Classification Task, Journal of Experimental Psychology, 83, 299-303.

Hogan, M. J., 1966, Influence of Motivation on Reactive Inhibition in Extraversion-Introversion, Perceptual and Motor Skills, 22, 187-192.

Holmes, D. L., K. Peper, L. W. Olsho, and D. E. Raney, 1978, Searching for Multiple Targets Simultaneously: Is it Really Possible, Perceptual and Motor Skills, 46, 227-234.

Hord, D. J., 1982, An EEG Predictor of Performance Decrement in a Vigilance Task, (Tech Report No. 82-?) San Diego, California: Naval Health Research Center.

Houghton, J. O., D. K. McBride, and K. Hannah, 1985, Performance and Physiological Effects of Acceleration-Induced (+Gz) Loss of Consciousness, Aviation, Space and Environmental Medicine, 56, 956-965.

Hoving, K. L., R. E. Morin, and D. S. Konick, 1970, Recognition Reaction Time and Size of the Memory Set: A Developmental Study, Psychonomic Science, 21, 247-248.

Howe, E. S., 1980, Effects of Partial Symmetry, Exposure Time, and Backward Masking on Judged Goodness and Reproduction of Visual Patterns, Quarterly Journal of Experimental Psychology, 32, 27-55.

Howe, E. S. and C. J. Brandau, 1983, The Temporal Course of Visual Pattern Encoding: Effects of Pattern Goodness, Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 35A, 607-633.

Hunter, D. R., 1975, Development of an Enlisted Psychomotor/Perceptual Test Battery, (AFHRL-TR-75-60), Air Force Human Resources Laboratory, Brooks Air Force Base, Texas.

Ichikawa, Shin-ichi, 1981, Rated Ease of Memorization of Dot-in-Matrix Patterns: Multiple Regression Analysis of Physical Variables which Describe Configuration of Dots, Japanese Psychological Research, 23, 69-78.

Jahnke, J. C., In Press, Error Factors in Recognition Memory, Proceedings of the 3rd Mid-Central Ergonomics/Human Factors Conference, Oxford, Ohio.

Jensen, A. R., 1965, Scoring the stroop test, Acta Psychologica, 24, 298-408.

Jensen, A. R. and W. D. Rohwer, Jr., 1966, The Stroop Color-Word Test: A Review, Acta Psychologica, 25, 36-93.

Jerison, H. H., C. W. Crannell, and D. Pownall, 1957, Acoustic Noise and Repeated Time Judgements in a Visual Movement Projection Task, (WADC-TR-57-54), Wright Air Development Center, Wright-Patterson Air Force Base, Ohio.

Jerison, H. J., 1977, Vigilance: Biology, Psychology, Theory, and Practice, In R. R. Mackie (Ed.), Vigilance: Theory, Operational Performance, and Physiological Correlates, New York: Plenum Press.

Jerison, H. J. and J. Arginteanu, 1958, Time Judgements, Acoustic Noise, and Judgement Drift, (WADC-TR-57-454) Wright Air Development Center, Wright-Patterson Air Force Base, Ohio (AD-130963).

Jex, H. R. and C. H. Cromwell, 1961, Theoretical and Experimental Investigation of some New Longitudinal Handling Quality Parameters, Aeronautical Systems Division, Wright-Patterson Air Fore Base, Ohio, TR 61-26.

Jex, H. R., McDonnell, and A. V. Phatak, 1966, A "Critical" Tracking Task for Manual Control Research, IEEE Transactions on Human Factors in Electronics, HFE-7, 138-144.

Jex, H. R., R. A. Peters, R. J. DiMarco, and R. W. Allen 1974, The Effects of Bedrest on Crew Performance During Simulated Shuttle Reentry, Vol. II, Control Task Performance, Washington, D.C.: NASA Contractor Report, NASA CR 2367.

Johnson, B. L. and W. K. Anger, 1983, Behavioral Toxicology, In W. Rom (Ed.) Environmental and Occupational Medicine, Boston: Little, Brown, and Co.

Johnson, B. L., H. H. Cohen, R. P. Struble, J. V. Setzer, W. K. Anger, B. D. Gutuik, T. McDonough, and P. Hauser, 1974, Field Evaluation of Carbon Monoxide Exposed Toll Collectors, In C. Xintaras and B. L. Johnson (Eds.), Behavioral Toxicology (306-328), U.S. Department HEW, Pub No. 74-126, U.S. Government Printing Office, Washington, D.C.

Johansen, G., C. Pfender, and W. Stein, 1976, Human Performance and Workload in Simulated Landing-Approaches with Autopilot-Failures, In T. B. Sheridan and G. Johansen (Eds.), Monitoring Behavior and Supervisory Control (83-95), New York: Plenum Press.

Kantowitz, B. H., S. G. Hart, and M. R. Bortolussi, 1983, Measuring Pilot Workload in a Moving-Base Simulator: I. Asynchronous Secondary Choice-Reaction Task, Proceedings of the 27th Annual Meeting of the Human Factors Society, Norfolk, VA, 319-322.

Kaplan, I. T. and T. Carvellas, T., 1965, Scanning for Multiple Targets, *Perceptual and Motor Skills*, 21, 239-243.

Keele, S. W., 1972, Attention Demands of Memory Retrieval, *Journal of Experimental Psychology*, 93, 245-248.

Keele, S. W. and H. L. Hawkins, 1982, Explorations of Individual Differences Relevant to High Level Skill, *Journal of Motor Behavior*, 14, 3-23.

Kennedy, R. S., W. P. Dunlap, M. B. Jones, N. E. Lane, and R. L. Wilkes, 1985, *Portable Human Assessment Battery: Stability, Reliability, Factor Structure, and Correlation with Tests of Intelligence*, Final Technical Report, National Science Foundation.

Kennedy, R. S., R. L. Wilkes, N. E. Lane, and J. L. Hamick, 1985, *Preliminary Evaluation of a Micro-Based Repeated Measures Testing System*, Essex Orlando Technical Report EOSR-85-1.

Ketchum, A. S., F. R. Sidell, E. B. Crowell, G. K. Ayhajanian, and A. H. Hayes, 1973, Atropine, Scopolamine, and Ditran: Comparative Pharmacology and Antagonists in Man, *Psychopharmacologia*, 28, 121-133.

Kirsner, K., J. E. Wells, and D. L. Sang, 1982, Physical and Name Match Processes in Same/Different Judgments for Letters, *Acta Psychologica*, 52, 73-85.

Klatzky, R. L. and Atkinson, 1971, Specialization of the Hemispheres in Scanning for Information in Short Term Memory, *Perception and Psychophysics*, 10, 335-338.

Klatzky, R. L., J. F. Juolea, and R. C. Atkinson, 1971, Test Stimulus Representation and Experimental Context Effects in Memory Scanning, *Journal of Experimental Psychology*, 87, 281-288.

Klein, G. S., 1964, Semantic Power Measured Through the Interference of Words with Color-Naming, *American Journal of Psychology*, 77, 576-588.

Klein, R. and R. Armitage, 1979, Rhythms in Human Performance: 1 1/2 Hour Oscillations in Cognitive Style, Science, 204, 1326-1328.

Klein, R. and H. R. Jex, 1975, Effects of Alcohol on a Critical Tracking Task, Journal of Studies on Alcohol, 36, 11-20.

Kosslyn, S. M. and S. P. Shwartz, 1977, A Simulation of Visual Imagery, Cognitive Science, 1, 265-295.

Krause, M. and A. C. Bittner, 1982, Repeated Measures on a Choice Reaction Time Task, Report No. NBDL-82R006, Naval Biodynamics Laboratory, New Orleans, Louisiana.

Kristofferson, J. W., 1972, Types and Frequency of Errors in Visual Search, Perception and Psychophysics, 11, 325-328.

Kristofferson, M. W., 1972a, Effects of Practice on Character Classification Performance, Canadian Journal of Psychology, 26, 540-560.

Kristofferson, M. W., 1972b, When Item-Recognition and Visual Search Functions are Similar, Perception and Psychophysics, 12, 379-384.

Kristofferson, M. W., M. Groen, and A. B. Kristofferson, 1973, When Visual Search Functions Look Like Item Recognition Functions, Perception and Psychophysics, 14, 186-192.

Krueger, L. E., 1978, A Theory of Perceptual Matching, Psychological Review, 85, 278-304.

Leonard, J. A., 1959, A 5-Choice Serial Reaction Apparatus, Applied Psychology Research Unit, Report No. 326/59, Cambridge, England.

Lewis, J., A. D. Baddeley, K. G. Bonham, and D. Lovett, 1970, Traffic Pollution and Mental Efficiency, Nature, 225, 95-96.

Lewis, V. J. and A. D. Baddeley, 1981, Cognitive Performance, Sleep Quality, and Mood During Deep Oxyhelium Diving, Ergonomics, 24, 773-793.

Lisher, B. J. and D. H. Glaister, 1978, The Effect of Acceleration and Seat Back Angle on Performance of a Reaction Time Task, FPRC Report No. 1364, Flying Personnel Research Committee, Ministry of Defense (Air Force Department), London.

Little, V. Z., B. O. Hartman, S. D. Leverett, 1968, Effects of Acceleration of Human Performance and Physiology with Special Reference to Transverse G, Aeromedical Review No. 4-68, USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC), Brooks Air Force Base, Texas.

Lively, B. L., 1972, Speed/Accuracy Tradeoff and Practice as Determinants of Stage Duration in a Memory-Search Task, Journal of Experimental Psychology, 96, 97-103.

Lively, B. L. and B. J. Sanford, 1972, The Use of Category Information in a Memory-Search Task, Journal of Experimental Psychology, 93, 379-385.

Logie, R. H. and A. D. Baddeley, 1983, A Trimix Saturation Dive to 660m: Studies of Cognitive Performance, Mood, and Sleep Quality, Ergonomics, 26, 359-374.

Lohman, D. F., 1979, Spatial Ability: A Review and Reanalysis of the Correlational Literature, (Tech Report No. 8), Stanford, California: Aptitude Research Project, School of Education, Stanford University.

Looper, M., 1976, The Effect of Attention Loading on the Inhibition of Choice Reaction Time to Visual Motion by Concurrent Rotary Motion, Perception and Psychophysics, 20, 80-84.

MacFlynn, G., E. A. Montgomery, G. W. Fenton, and W. Rutherford, 1984, Measurement of Reaction Time Following Minor Head Injury, Journal of Neurology, Neurosurgery, and Psychiatry, 47, 1326-1331.

Magiste, E., 1984, Stroop Tasks and Dichotic Translation: The Development of Interference Patterns in Bilinguals, Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 304-315.

Mandler, G., 1980, Recognizing: The Judgement of Previous Occurrence, Psychological Review, 87, 252-271.

Marmurek, H. H. C., 1977, Processing Letters in Words at Different Levels, Memory and Cognition, 5, 67-72.

Matarazzo, J. D., 1972, Wechsler's Measurement and Appraisal of Adult Intelligence, Baltimore: Williams and Wilkins.

McNair, D. M., R. J. Kahn, L. M. Frankenthaler, and L. L. Faldetta, 1984, Amoxapine and Amitriptyline, II, Specificity of Cognitive Effects During Brief Treatment of Depression, Psychopharmacology, 83, 134-139.

McRuer, D. T., D. Graham, E. Krendel, and W. Reisner, 1965, Human Pilot Dynamics in Compensatory Systems, (AFFDL-TR-65-15), U.S. Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio, (AD 470 337).

McRuer, D. T. and H. R. Jex, 1967, A Review of Quasi-Linear Pilot Models, IEEE Transactions on Human Factors in Electronics, 8, 231.

Michelson, M. J., 1961, Pharmacological Evidence of the Role of Acetylcholine in the Higher Nervous Activity of Man and Animals, Activas Nervosa Superior, 3, 140-147.

Michon, J. A., 1966, Tapping Regularity as a Measure of Perceptual Motor Load, Ergonomics, 9, 401-412.

Mihal, W. L. and G. V. Barrett, 1976, Individual Differences in Perceptual Information Processing and their Relation to Automobile Accident Involvement, Journal of Applied Psychology, 61, 229-233.

327

Mirsky, A. F. and P. V. Cardon, 1962, A Comparison of the Behavioral and Physiological Changes Accompanying Sleep Deprivation and Chlorpromazine Administration in Man, EEG and Clinical Neurophysiology, 14, 1-10.

Mohs, R. C., J. R. Tinklenberg, W. Y. Roth, and B. S. Kopell, 1980, Sensitivity of some Human Cognitive Functions to Effects of Methamephetamine and Secobarbital, Drug and Alcohol Dependence, 5, 145-150.

Moore, M. E. and B. M. Ross, 1963, Context Effects in Running Memory, Psychological Reports, 12, 451-465.

Moray, N., 1959, Attention in Dichotic Listening, Quarterly Journal of Experimental Psychology, 11, 56-60.

Moray, N., 1969, Listening and Attention, Baltimore: Penguin.

Morgan, B. B. and J. D. Repko, 1974, Evaluation of Behavioral Functions in Workers Exposed to Lead, In C. Xintaras, B. Johnson, and I. deGroot (Eds.), Behavioral Toxicology: Early Detection of Occupational Hazards, Washington, D.C.: U.S. Government Printing Office, U.S. DHHS (NIOSH) Publication No. 74-126.

Morin, R. E., D. V. DeRosa, and V. Stultz, 1967, Recognition Memory and Reaction Time, Acta Psychologica, 27, 298-305.

Morton, J., 1969, Categories of Interference: Verbal Mediation and Conflict in Card Sorting, British Journal of Psychology, 60, 329-346.

Naitoh, P., 1981, Circadian Cycles and Restorative Power of Naps, In L. C. Johnson, D. F. Tepas, W. P. Colquhoun, and M. J. Colligan (Eds.), Biological Rhythms, Sleep, and Shift Work, (553-580) New York: Spectrum.

Naus, J. J., S. Glucksberg, and P. A. Ornstein, 1972, Taxonomic Word Categories and Memory Search, Cognitive Psychology, 3, 643-654.

Naus, M. J., 1974, Memory Search of Categorized Lists: A Consideration of Alternative Self-Terminating Search Strategies, Journal of Experimental Psychology, 102, 992-1000.

Neisser, V., 1963, Decision-Time Without Reaction-Time: Experiments in Visual Scanning, American Journal of Psychology, 76, 376-385.

Neisser, V., R. Novick, and R. Lazar, 1963, Searching for Ten Targets Simultaneously, Perceptual and Motor Skills, 17, 955-961.

O'Reilly, J. P., 1977, Hana Kai II: A 17-Day Dry Saturation Dive at 18.6 ATA, VI: Cognitive Performance, Reaction Time, and Personality Changes, Undersea Biomedical Research, 4, 297-305.

Osborne, D. J. and Y. Rogers, 1983, Interactions of Alcohol and Caffeine on Human Reaction Time, Aviation Space and Environmental Medicine, 54, 528-534.

Ostfeld, A. M. and A. Aruguete, 1962, Central Nervous System Effects of Hyoscine in Man, Journal of Pharmacology and Experimental Therapeutics, 137, 133-139.

Pedhazur, E. J., 1982, Multiple Regression in Behavioral Research: Explanation and Prediction, New York: Holt, Rinehart, and Winston.

Pepper, R. L., R. S. Kennedy, A. C. Bittner, S. F. Wiker, and M. M. Harbeson, 1985, Performance Evaluation Tests for Environmental Research (PETER): Code Substitution Test, Perceptual and Motor Skills, 61, 735-745.

Perez, W. A., 1982, Mental Arithmetic: The Processing and Maintenance of Information in Working Memory, Unpublished Doctoral Dissertation, Miami University, Oxford, Ohio.

Perez, W. A., 1985, Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB) Procedures, Paper Presented at the Joint Working Group on Drug Dependent Degradation in Military Performance (JW6D3 MILPERF), Program Sub-Element In-Process Review, Frederic, Maryland.

329

Perez, W. A. and C. Tracy, 1983, Associative Confusion in Mental Addition and Subtraction, Paper Presented at the Annual Midwestern Psychological Association Meeting, Chicago, Illinois.

Pfefferbaum, A., J. M. Ford, W. T. Roth, and B. S. Kopell, 1980, Age Differences in P-300 Reaction Time Associations, Electroencephalography and Clinical Neurophysiology, 49, 257-265.

Phillips, W. A., 1974, On the Distinction Between Sensory Storage and Short Term Visual Memory, Perception and Psychophysics, 16, 283-290.

Pleban, R. J., D. A. Thomas, and H. L. Thompson, 1985, Behavior Research Methods, Instruments, and Computers, 17, 86-89.

Poltrock, S. E. and P. Brown, 1982, Individual Differences in Mental Imagery, AFOSR Tech Report 82-0149, Bolling Air Force Base, Washington, D.C.: Air Force Office of Scientific Research.

Posner, M. I. and S. Boies, 1971, Components of Attention, Psychological Review, 78, 391-408.

Posner, M. I., S. J. Boies, W. H. Eichelman, and R. Taylor, 1969), Retention of Visual and Name Codes of Single Letters, Journal of Experimental Psychology Monograph, 79, 1-16.

Posner, M. I. and R. F. Mitchell, 1967, Chronometric Analysis of Classifications, Psychological Review, 74, 392-409.

Poulton, E. C., G. M. Hunt, A. Carpenter, and R. S. Edwards, 1978, The Performance of Junior Hospital Doctors Following Reduced Sleep and Long Hours of Work, Ergonomics, 21, 279-295.

Pritchard, W. S., 1981, Psychophysiology of P300, Psychological Bulletin, 89, 506-540.

Putz. V. R., 1979, The Effects of Carbon Monoxide on Dual Task Performance, Human Factors, 21, 13-24.

Reader, D. C., R. A. Benel, and A. J. Rahe, 1981, Evaluation of a Manikin Psychomotor Task (Report SAM-TR-81-10), Brooks Air Force Base, Texas: USAF School of Aerospace Medicine.

Repko, J. D., P. E. Jones, L. S. Garcia, E. J. Schneider, E. Roseman, and C. R. Corum, 1976, Behavioral and Neurological Effects of Methyl Chloride, Washington, D.C.: U.S. Government Printing Office, U.S. DHHS (NIOSH) Publication No. 77-125.

Repko, J. D., B. B. Morgan, Jr., and J. Nicholson, 1975, Behavioral Effects of Occupational Exposure to Lead, DHEW Publication No. (NIOSH) 75-184, Washington, D.C.: U.S. Government Printing Office.

Restle, F., 1970, Speed of Adding and Comparing Numbers, Journal of Experimental Psychology, 83, 274-278.

Rosa, R. R., D. D. Wheeler, J. S. Warm, and M. J. Colligan, 1985, Extended Workdays: Effects on Performance and Ratings of Fatigue and Alertness, Behavior Research Methods, Instruments and Computers, 16, 256-261.

Ross, B. M., 1966a, Serial-Order as a Unique Source of Error in Running Memory, Perceptual and Motor Skills, 23, 195-209.

Ross, B. M. 1966b, Serial-Order Effects in Two-Channel Running Memory, Perceptual and Motor Skills, 23, 1099-1107.

Ross, B. M. and R. M. Chambers, 1967, Effects of Transvers G-Stress on Running Memory, Perceptual and Motor Skills, 24, 423-435.

Ross, J., 1970, Extend Practice with a Single-Character Classification Task, Perception and Psychopysics, 8, 276-278.

Rosvald, H. E., A. F. Mirsky, I. Sarason, E. D. Bransome, and L. N. Beck, 1956, A Continuous Performance Test of Brain Damage, Journal of Consulting Psychology, 90, 343-350.

Roth, W. T., J. R. Tinklenberg, and B. S. Kopell, 1977, Ethanol and Marihuana Effects on Event-Related Potentials in a Memory Retrieval Paradigm, Electroencephalography and Clinical Neurophysiology, 42, 381-388.

Royer, F. L., 1981, Detection of Symmerty, Journal of Experimental Psychology: Human Perception and Performance, 7, 1186-1210.

Ryman, D. M., P. Naitoh, and C. E. Englund, 1984, Minicomputer Administered Tasks in the Study of Effects of Sustained Work on Human Performance, Behavior Research Methods, Instruments, and Computers, 16, 256-261.

Sax, D. S., B. O'Donnell, N. Butters, L. Menzer, K. Montgomery, and H. L. Kayne, 1933, Computed Tomographic, Neurologic, and Neurophysiological Correlates of Huntington's Disease, International Journal of Neuroscience, 18, 21-36.

Schnore, M. M. and J. T. Partington, 1967, Immediate Memory of Visual Patterns: Symmetry and Amount of Information, Psychonomic Science, 8, 421-422.

Schwent, V. L., E. Snyder, and S. A. Hillyard, 1976, Auditory Evoked Potentials During Multichannel Selective Listening: Role of Pitch and Localization Cues, Journal of Experimental Psychology: Human Perception and Performance, 2, 313-325.

Seales, D. M., R. S. Kennedy, and A. C. Bittner, Jr., 1980, Development of Performance Evaluation Tests for Environmental Research (PETER): Arithmetic Computation, Perceptual and Motor Skills, 51, 1023-1031.

Seppala, T. and R. Visakorpi, 1983, Psychophysiological Measurements after Oral Atropine in Man, Acta Pharmacologica et Toxicologica, 52, 68-74.

Shepard, R. N. and J. Metzler, 1971, Mental Rotation of Three-Dimensional Objects, Science, 171, 701-703.

Shingledecker, C. A., 1980, Serial Production of Brief Temporal Intervals as an Index of Operator Workload: The Effects of Tracking Task Demand on Performance Regularity in Manual and Pedal Response Modes, Technical Report SRL-TSG-TR-80-08, Systems Research Laboratories, Inc., Dayton, Ohio.

Shingledecker, C. A., 1984, A Task Battery for Applied Human Performance Assessment Research, Technical Report No. AFAMRL-TR-84-071, Air Force Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio.

Shingledecker, C. A., W. H. Acton, and M. S. Crabtree, 1983, Development and Application of a Criterion Task Set for Workload Metric Evaluation, Paper Presented at the Second Aerospace Behavioral Engineering Technology Conference, (No. 83/419), Long Beach, California.

Slobin, D. I., 1966, Grammatical Transformations and Sentence Comprehension in Childhood and Adulthood, Journal of Verbal Learning and Verbal Behavior, 5, 219-227.

Smith, E. E., 1968, Choice Reaction Time: An Analysis of the Major Theoretical Positions, Psychological Bulletin, 69, 77-110.

Smith, P. J and G. D. Langolf, 1981, The Use of Sternberg's Memory-Scanning Paradigm in Assessing Effects of Chemical Exposure, Human Factors, 23, 701-708.

Spicuzza, R. J., A. R. Pinkus, and R. D. O'Donnell, 1974, Development of Performance Assessment Methodology for the Digital Avionics Information System, Final Report, USAF Aerospace Medical Research Laboratory, Dayton, Ohio, 36-54.

Stazyk, E. H., M. H. Ashcraft, and M. S. Hamann, 1982, A Network Approach to Mental Multiplication, Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 320-335.

Sternberg, S., 1966, High Speed Scanning in Human Memory. Science, 153, 652-654.

Sternberg, S., 1967, Two Operators in Character Recognition: Some Evidence from Reaction-Time Measurements, Perception and Psychophysics, 2, 45-53.

Sternberg, S., 1969a, Memory-Scanning: Mental Processes Revealed by Reaction-Time Experiments, American Scientist, 57, 421-457.

Sternberg, S., 1969b, The Discovery of Processing Stages: Extensions of Donders' Method, Acta Psychologica, 30, 276-315.

Sternberg, S., 1975, Memory-Scanning: New Findings and Current Controversies, Quarterly Journal of Experimental Psychology, 27, 1-32.

Stroop, J. R., 1935, Studies of Interference in Verbal Reactions, Journal of Experimental Psychology, 18, 643-662.

Sverko, B., 1977, Individual Differences in Time-Sharing Performance, Technical Report ARL-77-4/AFOSR-77-4, Aviation Research Laboratory.

Swanson, J. M., A. M. Johnsen, and G. E. Briggs, 1972, Recoding in a Memory Search Task, Journal of Experimental Psychology, 93, 1-9.

Sykes, D. H., V. I. Douglas, and G. Morgenstern, 1973, Sustained Attention in Hyperactive Children, Journal of Child Psychology and Child Psychiatry, 14, 213-221.

Taub, J. M., 1982, Effects of Scheduled Afternoon Naps and Bedrest on Daytime Alertness, International Journal of Neuroscience, 16, 107-127.

The Military Performance Working Group, 1983, The Effect of Chemical Warfare Treatment Drugs on Military Performance: A Tri-Service Drug Screening Program, Working Paper, U.S. Army Research and Development Command, Fort Detrick, Maryland.

334

Theios, J., P. G. Smith, S. E. Haviland, J. Traupmann, and M. C. Moy, 1973, Memory Scanning as a Serial Self-Terminating Srocess, Journal of Experimental Psychology, 97, 323-336.

Thorson, G., L. Hochhaus, and R. F. Stanners, 1976, Temporal Cnanges in Visual and Acoustic Codes in a Letter-Matching Task, Perception and Psychophysics, 19, 346-348.

Tilley, A. J., R. T. Wilkinson, P. S. G. Warren, B. Watson, and M. Drud, 1982, The Sleep and Performance of Shift Workers, Human Factors, 24, 629-641.

Tone, T., 1981, Serial Checking of Letter Targets, Perceptual and Motor Skills, 53, 291-297.

Townsend, J. T., 1971, A Note on the Identifiability of Parallel and Serial Processes, Perception and Psychophysics, 10, 161-163.

Townsend, P. E, and L. C. Johnson, 1979, Relation of Frequency Analyzed EEG to Monitoring Behavior, EEG and Clinical Neurophysiology, 47, 272-279.

Treisman, A., 1960, Contextual Cues in Selective Listening, Quarterly Journal of Experimental Psychology, 12, 242-248.

Treisman, A., 1964, The Effect of Irrelevant Material on the Efficiency of Selective Listening, American Journal of Psychology, 77, 533-546.

Treisman, A. M., 1969, Strategies and Models of Selective Attention, Psychological Review, 76, 282-299.

Tuttle, T. C., G. D. Wood, and C. B. Grether, 1976, Behavioral and Neurological Evaluation of Workers Exposed to Carbon Disulfide, Washington, D.C.: U.S. Government Printing Office, U.S. DHHS (NIOSH) Publication No. 77-214.

Tuttle, T. C., G. D. Wood, C. B. Grether, B. L. Johnson, and C. Xintaras, 1977, A Behavioral and Neurological Evaluation of Dry Cleaners Exposed to Perchlorethylene, Washington, D.C.: U.S. Government Printing Office, U.S. DHHS (NIOSH) Publication No. 77-214.

Underwood, B. J., 1983, Attributes of Memory, Glenview, Illinois: Scott Foresman.

Van Patten, R. E., 1984, Tolerance, Fatigue, Physiological, and Performance Effects of Sustained and Oscillating Lateral Acceleration, In AGARD Conference Proceedings (No. 371) on Human Factors Considerations in High Performance Aircraft, Reference No. 7.

Vidulich, M. and C. D. Wickens, 1981, Time-Sharing Manual Control and Memory Search: The Joint Effects of Input and Output Modality Competition, Priorites and Control Order, Tech Report No. EPL-81-4, Department of Psychology, University of Illinois, Champaign, Illinois.

Vroon, P. A., 1973, Tapping Rate as a Measure of Expectancy in Terms of Response and Attention Limitation, Journal of Experimental Psychology, 101, 183-185.

Vroon, P. A. and A. G. Vroon, 1973, Tapping Rate and Expectancy in Simple Reaction Time Tasks, Journal of Experimental Psychology, 98, 85-90.

Wanner, E. and S. Shiner, 1976, Measuring Transient Memory Load, Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 320-335.

Warm, J. S., R. G. Wait, and M. Loeb, 1976, Head Restraint Enhances Visual Monitoring Performance, Perception and Psychophysics, 20, 299-304.

Warren, L. R. and G. R. Marsh, 1979, Changes in Event Related Potentials During Processing of Stroop Stimuli, International Journal of Neuroscience, 9, 217-223.

Wason, P. C., 1961, Response to Affirmative and Negative Binary Statements, British Journal of Psychology, 52, 133-142.

Wattenbarger, B. L., 1968, Speed and Accuracy Set in Visual Search Performance, Paper Presented at the Meeting of the Midwestern Psychological Association, Chicago, Illinois.

Webb, W. B., 1985, Experiments on Extended Performance; Repetition, Age, and Limited Sleep Periods, Behavior Research Methods, Instruments, and Computers, 17, 27-36.

Webb, W. B. and C. M. Levy, 1984, Effects of Spaced and Repeated Total Sleep Deprivation, Ergonomics, 27, 45-58.

Wechsler, D., 1958, The Measurement and Appraisal of Adult Intelligence, Baltimore: The Williams and Wilkins Co.

Wetherell, A., 1981, The Efficacy of Some Auditory--Vocal Subsidiary Tasks as Measures of the Mental Load on Male and Female Drivers, Ergonomics, 24, 197-114.

Wickens, C. D., 1976, The Effects of Divided Attention on Information Processing in Manual Tracking, Journal of Experimental Psychology: Human Perception and Performance, 2, 1-13.

Wickens, C. D., 1981, Processing Resources in Attention, Dual Task Performance, and Workload Assessment, Technical Report EPL-81-3, Engineering Psychology Research Laboratory, University of Illinois at Urbana-Champaign.

Wickens, C. D., 1984, Engineering Psychology and Human Performance, Columbus, Ohio: Charles E. Merrill Publishing Co.

Wickens, C. D., S. J. Mountford, and W. Schreiner, 1980, Task Dependent Differences and Individual Differences in Dual Task Performance, Report No. NBDL-M003, Naval Biodynamics Laboratory, New Orleans, Louisiana.

337

Wickens, C. D. and D. Sandry, 1982, Task-Hemispheric Integrity in Dual Task Performance, Acta Psychologica, 52, 227-247.

Wickens, C. D., D. L. Sandy, and R. Hightower, 1982, Display Location of Verbal and Spatial Material: The Joint Effects of Task-Hemispheric Integrity and Processing Strategy, Tech Report No. EPL-82-2, Department of Psychology, University of Illinois, Champaign, Illinois.

Wickens, C. D., M. Vidulich, D. Sandy, and S. Schiflett, 1981, Factors Influencing the Performance Advantage of Speech Technology, Proceedings, 25th Annual Meeting of the Human Factors Society.

Wiker, S. F., R. S. Kennedy, M. E. McCauley, and R. L. Pepper, 1979, Susceptibility to Seasickness: Influence of Hull Design and Steaming Direction, Aviation, Space, and Environmental Medicine, 50, 1046-1051.

Wiker, S. F. and R. L. Pepper, 1978, Change in Crew Performance, Physiology, and Effective State Due to Motions Aboard a Small Monohull Vessel: A Preliminary Study, Washington, D.C.: Department of Transportation, (Technical Report No. CG-D-75-78).

Wilkinson, R. T. and D. Houghton, 1975, Portable Four-Choice Reaction Time Test with Magnetic Memory, Behavior Research Methods and Instrumentation, 7, 441-446.

Wingfield, A. and A. A. Branca, 1970, Strategy in High-Speed Memory Search, Journal of Experimental Psychology, 83, 63-67.

Winkelman, J. H. and J. Schmidt, 1974, Associative Confusion in Mental Arithmetic, Journal of Experimental Psychology, 102, 734-736.

Xintaras, C., J. R. Burg, S. Tanaka, S. T. Lee, B. L. Johnson, C. A. Cottrill, and J. Bender, 1978, NIOSH Health Survey of Velsical Pesticide Workers, Washington, D.C.: U.S. Government Printing Office, U.S. DHHS (NIOSH) Publication No. 78-136.